Structural Studies of Large N-terminal Domains of Y Receptors

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Chao Zou

aus

P. R. China

Promotionskomitee

Prof. Dr. Oliver Zerbe (Vorsitz und Leitung der Dissertation)

Prof. Dr. John Robinson

Zürich, 2009

Table of Contents

Abbreviation 1

Summary 3

Zusamenfassung 6

Chapter 1

Introduction

1.1 G-protein coupled receptors (GPCRs) 9

1.1.1 Classification of GPCRs 10

1.1.2 Current state of structural studies of GPCRs 12

1.1.3 Oligomerization of GPCRs 15

1.1.4 Posttranslational modifications of GPCRs 16

1.2 Neuropeptide Y receptors (Y receptors) 20

1.2.1 Evolution and members of Y receptors 20

1.2.2 Physiological functions of Y receptors 21

1.2.3 Mutagenesis studies of Y receptors 24

1.2.4 Which portions of Y receptors are important for ligand binding? 26

1.3 The ligands for Y receptors: neuropeptide Y (NPY) hormones 29

1.3.1 Members of NPY family 29

1.3.2 Structural studies of NPY members 30

1.3.3 A novel binding model of NPY hormones with Y receptors 33

1.4 Using protein fragments to study membrane protein structure 35

1.4.1 The 2-step membrane protein folding model 35

1.4.2 Current progress in studies of transmembrane domains 36

1.4.3 Membrane mimetics for NMR studies 39

1.5 Production of membrane protein fragments 44

1.5.1 The fusion method 44

1.5.2 Direct expression 45

1.5.3 Strain selection 45

1.5.4 Conditions for culturing 46

1.5.5 Potential problems 47

1.6 Summary of the work described in this thesis 48

1.7 References 52

Chapter 2

Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family

2.1 Introduction 66

2.2 Results 69

2.2.1 Recombinant production of N-Y4 69

2.2.2 The structure of N-Y4 70

2.2.3 Topology of membrane-association 72

2.2.4 Immobilizing the N terminus on the membrane 74

2.2.5 Interaction between N-Y4 and neuropeptides from the NPY family 74

2.3 Discussions 77

2.4 Conclusions 81

2.5 Materials and Methods 82

2.5.1 Plasmid Construction, Expression and Purification of N-Y4 82

2.5.2 Synthesis and purification of the neuropeptides and of unlabelled N-terminal fragments 83

2.5.3 Dodecylphosphoethanolamine Coupling to Carboxyl Terminus of N-Y4 84

2.5.4 NMR experiments 84

2.5.5 Membrane-association topology using spin labels 85

2.5.6 Surface Plasmon Resonance (SPR) studies 85

2.6 Acknowledgement 86

2.7 References 87

2.8 Supplementary Materials 91

Chapter 3

Properties of the N-terminal domains from Y receptors probed by NMR spectroscopy

3.1 Introduction 100

3.2 Results 102

3.2.1 Expression of N-terminal domains in isotopically-labeled form 102

3.2.2 Assignment of chemical shifts 104

3.2.3 Screening structural properties by ^{15}N relaxation and CD spectroscopy 105
3.2.4 Structures of the N-terminal domains in presence of micelles 107
3.2.5 Interaction studies with neuropeptides from the NPY family 108
3.3 Discussions 112
3.4 Materials and methods 115
3.4.1 Materials 115
3.4.2 Expression and purification of N-terminal domains 115
3.4.3 NMR spectroscopy 117
3.5 Acknowledgements 118
3.6 References 119
3.7 Supplementary materials 122

Chapter 4

Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a Human GPCR

4.1 Introduction 133
4.2 Results 137
4.2.1 Optimization of protein expression 137
4.2.2 Optimization of purification and detergent 138
4.2.3 Spectroscopy and backbone assignment 140
4.2.4 Secondary structure 142
4.3 Discussions 145
4.4 Conclusions 149
4.5 Materials and methods 150
4.5.1 Plasmid construction 150
4.5.2 Protein expression and purification 150
4.5.3 NMR sample preparation 151
4.5.4 NMR spectroscopy and backbone assignment 151
4.5.5 Circular Dichroism spectroscopy 152
4.6 Acknowledgements 152
4.7 References 153
4.8 Supplementary materials 160

Chapter 5

Recognition of neurohormones of the NPY family by their receptors

5.1 Introduction 164
5.2 Structural features of peptides from the neuropeptide Y family in solution 167
5.3 Structural features of peptides from the neuropeptide Y family in the membrane-bound state 170
5.4 A series of events of receptor binding 173
5.5 Investigating the structural transition between bulk solution and the membrane environment 176
5.6 Removal of the back-fold when diffusing towards the membrane occurs cooperatively 178
5.7 Weak interactions between the hormones and the N-terminal domains are observed 180
5.8 Outlook 182
5.9 Acknowledgements 183
5.10 References 184
Curriculum Vitae 188
Publications and conferences 189
Acknowledgements 190

Abbreviation

AM	adrenomedullin
AR	adrenergic receptor
CCR2	chemokine C-C motif receptor 2
CCR5	chemokine C-C motif receptor 5
CGRP	calcitonin generelated peptide
CMC	critical micelle concentration
CNS	central nervous system
CX3C	chemokine C-X3-C motif receptor
DCM	dichloromethane
DDM	dodecyl-β-d-maltoside
DHPC	1,2-Diheptanoyl-sn-Glycero-3-Phosphocholine
DIEA	diisopropylethylamine
DMF	dimethyl formamide
DPC	dodecylphosphocholine
Doxyl	(4,4-dimethyl-3-oxazolidine-N-oxyl)
DTT	1,4-dithiothreitol
ECL	extracellular loop
EDTA	ethylenediamine tetraacetic acid
ELISA	enzyme linked immunostaining assay
ESI-MS	electrospray ionization mass spectroscopy
Fmoc	9-fluorenylmethoxycarbonyl
GABA	gamma amino butyric acid
GdnHCL	guanidinium hydrochloride
GPCR	G-protein coupled receptor
GST	glutathione S transferase
HATU	2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	2-(1H-benzotriazol-1-yL)-1,1,3,3-tetramethyluronium-hexafluorophosphat)
HEPES	2-(4-(2-hydroxyethyl)- 1-piperazinyl)-ethanesulfonic acid
HOBT	N-hydroxybenzotriazole
HSQC	heteronuclear single-quantum correlation

ICL	intracellular loop
IPTG	isopropyl β-D-1-thiogaloctopyranoside
K_D	dissociation constant
KSI	ketosteroidisomerase
LDAO	Lauryldimethylamine-oxide
LMPG	1-Myristoyl-2-Hydroxy-sn-Glycero-3-[Phospho- rac-(1-glycerol)]
LPPG	1-Palmitoyl-2-Hydroxy-sn-Glycero-3-[Phospho- rac-(1-glycerol)]
MALDI-TOF	matrix-assisted Laser desorption ionization-time of flight
MBP	maltose binding protein
MCP-1	monocyte chemoattractant protein 1
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
NPY	neuropeptide (h, human; p, porcine)
N-Y4	N-terminal domain of Y4 receptor
Ni-NTA	nickel-nitrilotriacetic acid
OGP	octyl gluco pyranoside
PAM	peptidylglycine α-amidating monooxygenase
PP	pancreatic polypeptide (a, avian; b, bovine; h, human)
PYY	peptide YY
RP-HPLC	reverse phase-high performance liquid chromatography
RT	room temperature
RU	response units; arbitrary unit in SPR
SDS	sodium dodecyl sulfate
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
TEV	tobacco etch virus
TFA	trifluoroacetic acid
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimized spectroscopy
TM	transmembrane

Summary

During my Ph.D. study I have investigated two principal aspects of GPCR structural biology in two stages: In the first stage I have focused my attention on structural studies of N termini of all the Y receptors and their interactions with natural ligands from neurohormone Y family, mainly by using NMR as well as by other biochemical and biophysical methods; in the second stage I extended my work on the N terminus of the Y4 receptor to a longer construct which comprises the N terminus, the first transmembrane domain, the first intracellular loop, the second transmembrane domain and the first extracellular loop. This fragment is much longer and thus more native-like. With this construct I hope to more accurately shed light on the structure and function of the entire Y4 receptor.

In the first stage we (the work of N terminus of Y2 receptor was conducted by my colleague Reto Walser) have expressed and purified by recombinant methods all the four N terminal domains from Y receptors. In case of N-Y2 and N-Y5, a soluble ubiquitin fusion protein was used to facilitate overexpression and purification and the target peptides were liberated later with help of the yeast ubiquitin hydrolase. In case of N-Y1 and N-Y4, unspecific cleavages were observed when the soluble fusion method was employed. To avoid unspecific cleavage, we produced the target protein in inclusion bodies by utilizing ketosteroidisomerase as fusion partner. Since there are a number of methionine residues in the target sequence, the classical chemical cleavage method (CNBr) could not be applied, and enzymatic cleavage is the only option. Extensive detergent screening indicated that sarcosyl could be used for that purpose, an anionic detergent, which can solubilize inclusion bodies and maintain the activity of the used TEV protease. In general, milligram quantities of labeled purified peptides can be obtained by either the soluble or insoluble fusion method from 1L of culture, which is sufficient for the following NMR studies. The interactions between the N termini of Y receptors and the neuropeptides of the NPY family are studied by chemical shift mapping methods using ^{15}N labeled neuropeptides and surface-plasmon resoance spectroscopy (BIA Core). The chemical shift mapping data demonstrated weak binding between the peptides corresponding to the N termini and the neuropeptides. This result is consistent with a model for binding of peptides from the NPY family to their GPCRs that postulates membrane association prior to receptor binding. The neuropeptide residues with large chemical shift changes display

surprising conservation in sequence despite much variability for other residues of the N termini. Two regions in the neuropeptides, which are mostly affected by the N termini, are the hinge region and the α helix in the C terminus. Chemical shift mapping of the neuropeptides upon addition of N-Y1 and N-Y4 demonstrate high similarity, which is consistent with their common origin as detected in phylogenetic studies. BIA Core measurements suggested that there is a specific binding between N-Y4 and PP and the binding affinity is around 50μM. Further mutagenesis experiments pinpointed that the positively charged residues from N-Y4 and that negatively charged residues from PP are responsible for this binding. The presence of specific binding between N-Y4 and PP indicates that the N terminal domain from the Y4 receptor might not only transiently bind PP and thus help transferring the ligand into the real binding pocket but may also provide specificity for various ligands. Moreover, binding of the N termini to DPC micelles results in larger changes of structure and dynamics, once more highlighting the significance of membrane association during ligand recognition and binding. Importantly in case of N-Y4, α helices have been observed in both the N and C terminal parts in presence of DPC micelles, whereas the peptide is completely unstructured in the absence of the DPC micelles. The structure of N-Y4 does not depend on the presence of the first tramsmembrane domain, instead the membrane surface alone is sufficient to induce secondary structure.

In the second stage on basis of structure and functional studies of N-Y4, I have extended my research to a longer construct, which contains more domains from the entire receptor, thus better mimicking the native properties in principle. For the expression of this protein, a direct expression without fusion to another protein was used. The non-fusion method bypasses the chemical or enzymatic cleavage of fusion proteins, hence simplifying the whole purification and increasing the final yield. After systematic optimization of cell strains, induction temperature and inducer concentration, optimal conditions were found when using the BL21-AI strain and 0.2% arabinose at 20^{o}C for induction. Under such conditions around 6 mg labeled protein can be purified from 1L of culture as fast as within 3 days. Sample purity and homogeneity are critical for a perfect NMR spectrum, and during the purification of this hydrophobic protein two steps are found indispensable: (1) the presence of multiple cysteine residues lead to aggregation even in 1mM merceptoethanol,

therefore addition of high concentration of reducing agents was required: 100mM DTT and 250mM merceptoethanol is essential to remove all aggregates and facilitates the subsequent HPLC purification; (2) HPLC is necessary to remove non-proteineous contaminants, which are invisible on SDS-PAGE. Membrane proteins can only exert their functions when they are appropriately integrated in the membrane, therefore detergent-formed micelles are employed here as a membrane-mimetic. Failure to find proper detergent micelles encouraged me to explore mixtures of detergents, and 1% DPC/6% LPPG proves to be the best for mimicking membrane environment for this protein. With ^{2}H, ^{13}C and ^{15}N labeled protein, the backbone can be assigned almost completely. The CD spectrum and the chemical shift of α and β carbons indicate a predominant alpha helical conformation. The sidechain assignment and tertiary structure calculation is undergoing presently.

Zusammenfassung

Während meiner Doktorarbeit habe ich zwei wichtige Aspekte der GPCR Strukturbiologie untersucht: In der ersten Phase habe ich meine Aufmerksamkeit auf strukturelle Studien von N-Termini aller Y-Rezeptoren und deren Wechselwirkungen mit natürlichen Liganden der Neuropeptid Y Familie gerichtet; insbesondere durch Nutzung der NMR Spektroskopie sowie anderer biochemischer und biophysikalischer Methoden. In der zweiten Phase habe ich meine Arbeit auf ein längeres Konstrukt gerichtet, das neben dem N-Terminus des Y4-Rezeptors die erste Transmembran-Domäne, den ersten intrazellulären Loop, die zweite Transmembran-Domäne und die erste extrazelluläre Schleife enthält. Dieses Fragment ist wesentlich länger und damit dem Gesamtrezeptor ähnlicher. Mit diesem Konstrukt hoffe ich, dass ich genauere Aussagen über die Struktur und Funktion des gesamten Y4-Rezeptor machen kann.

Anfänglich (die Arbeit über den N-Terminus des Y2-Rezeptor wurde von meinem Kollegen Reto Walser durchgeführt) habe ich alle vier N-terminalen Domänen der Y-Rezeptoren mittels rekombinanter Methoden exprimiert und aufgereinigt. Im Falle des N-Y2 und N-Y5 wurde ein lösliches Ubiquitin-Fusionsprotein eingesetzt, dass die Überexpression und Aufreinigung erleichtert. Die Ziel-Peptide wurden später mit Hilfe der Ubiquitinhydrolyase abgespalten. Im Falle des N-Y1 und N-Y4 wurden unspezifische Spaltungen beobachtet, als die lösliche Fusionsprotein-Methode benutzt wurden. Um die unspezifische Spaltung zu vermeiden, haben wir das Zielprotein in "Einschlusskörpern" produziert durch Einsatz der Ketosteroidisomerase als Fusionspartner. Weil es mehrere Methionine in der Zielsequenz gibt, konnte die klassische chemische Spaltungsmethode (CNBr) nicht angewandt werden, und die enzymatische Spaltung war die einzige Option. Umfangreiches Detergenzien-Screening zeigte, dass Sarcosyl, ein anionisches Detergenz, verwendet werden kann. Dieses vermag die "Einschlusskörper" zu solubilisieren, und erhält die Aktivität der verwendeten TEV Protease. In der Regel können Milligramm-Mengen von markiertem Peptide entweder mit der löslichen oder der unlöslichen Fusionsmethode aus einem Liter Kultur gereinigt werden. Die Menge ist ausreichend für die folgenden NMR-Studien. Die Wechselwirkungen zwischen den N-Termini der Y-Rezeptoren und den Neuropeptiden der NPY-Familie wurden durch 'chemical shift mapping' unter Einsatz von ^{15}N markierten Neuropeptiden und 'surface plasmon resoance'-Spektroskopie (BIA-Core) untersucht. Die 'chemical shift mapping' Daten zeigen

schwache Bindung zwischen den N-Termini und den Neuropeptiden. Dieses Ergebnis steht im Einklang mit einem Modell für die Bindung zwischen Peptiden aus der NPY Familie und ihren GPCRs, das die Membranassoziation als ein der Rezeptorbindung vorgelagerten Schritt postuliert. Die Neuropeptid-Reste mit großen 'chemical shift' Änderungen zeigten überraschende Sequenzkonservation trotz erheblicher Variabilität bei anderen Resten im N Terminus. Zwei Regionen in den Neuropeptiden, die vor allem Kontakte mit den N-Termini machen, sind die Scharnier-Region und die C-terminale α-Helix. Resultate der 'Chemical shift mapping' Experimente der Neuropeptide nach Zusatz von N-Y1 und N-Y4 zeigen grosse Ähnlichkeit, was im Einklang mit ihrer gemeinsamen Herkunft, die in phylogenetischen Studien erkundet wurde, steht. BIA-Core-Messungen legen nahe, dass es eine spezifische Bindung zwischen N-Y4 und dem pankreatischen Polypeptid (PP) gibt, und die Affinität ist rund 50 μM. Weitere Mutagenese Experimente ermittelten, dass die positiv geladenen Reste aus N-Y4 und die negativ geladenen Reste aus PP für diese Bindung verantwortlich sind. Das Vorhandensein von spezifischer Bindung zwischen N-Y4 und PP weist darauf hin, dass die N-terminale Domäne des Y4-Rezeptor nicht nur vorübergehend PP bindet und dabei hilft, den Liganden in die wahre Bindetasche zu übertragen, sondern auch zur Spezifität der Lingandenerkennung beiträgt. Darüber hinaus resultiert die Bindung zwischen N-Termini und DPC Mizellen in größeren Veränderungen von Struktur und Dynamik; dies weist noch einmal auf die Bedeutung der Membranassoziation bei der Ligandenerkennung und – bindung hin. Im Fall von N-Y4 wurden in Anwesenheit von DPC Mizellen α-Helices in den N- und C-terminalen Segmenten beobachtet, während das Peptid in Abwesenheit der DPC Mizellen völlig unstrukturiert ist. Diese Struktur von N-Y4 hängt nicht von der Anwesenheit der ersten Transmembran-Domäne ab, stattdessen ist die Membran-Oberfläche allein ausreichend, um die Sekundärstruktur zu induzieren.

In der zweiten Phase, auf der Grundlage von strukturellen und funktionellen Studien von N-Y4 habe ich mein Augenmerk auf ein längeres Konstrukt erweitert, das mehrere Domänen aus dem Gesamt-Rezeptor enthält, daher prinzipiell besser die nativen Eigenschaften imitiert. Für die Expression dieses Proteins wurde eine direkte Expression ohne Fusion zu einem anderen Protein verwendet. Die Nicht-Fusionsmethode umgeht die chemische oder enzymatische Spaltung des Fusionsproteins, daher vereinfacht sie die gesamte Reinigung und erhöht die endgültige Ausbeute. Nach der systematischen Optimierung der Zellstämme,

Induktionstemperatur und Induktor-Konzentration, wurden optimalen Bedingungen gefunden bei der Verwendung des BL21-AI Stamm und 0,2% Arabinose bei 20^{o}C für die Induktion. Unter diesen Bedingungen kann etwa 6 mg gereinigtes und markiertes Protein aus einem Liter Kultur innerhalb von 3 Tagen hergestellt werden. Probenreinheit und Homogenität sind entscheidend für ein befriedigendes NMR-Spektrum. Während der Reinigung von diesem hydrophoben Protein sind zwei Schritte unverzichtbar: (1) das Vorhandensein von mehreren Cystein-Resten führt zur Aggregation, auch in 1 mM Merceptoethanol; daher war der Zusatz von hohen Konzentrationen an Reduktionsmittel erforderlich: 100 mM DTT und 250 mM Merceptoethanol ist von wesentlicher Bedeutung, um alle Aggregate zu reduzieren und die spätere Reinigung via HPLC zu erleichtern; (2) HPLC ist notwendig zur Beseitigung von nicht-Protein Kontaminationen, die nicht sichtbar auf dem SDS-PAGE sind. Membranproteine können nur dann ihre Funktionen ausüben, wenn sie richtig in die Membran integriert sind, daher werden Detergenz-Mizellen hier als Membran-Mimetika verwendet. Da anfänglich keine geeigenten Detergenzien gefunden werden konnten, habe ich Detergenz Mischungen auszuprobiert; und 1% DPC / 6% LPPG erweist sich als die beste Mischung, um die Membran-Umgebung für dieses Protein zu imitieren. Mit ^{2}H, ^{13}C und ^{15}N markiertem Protein konnte das Rückgrat fast komplett zugeordnet werden. Das CD-Spektrum und die chemischen Verschiebungen der α-und β-Kohlenstoffe deuten auf ein predominant α –helicale Struktur hin. Die Seitenketten-Zuordnung und Berechnung der Tertiärstruktur erfolgt derzeit.

Chapter 1

Introduction

Membrane proteins (MPs), also known as integral membrane proteins (IMPs), are the most abundant proteins in various prokaryotic and eukaryotic organisms and can account for 20-30% of the total genome[1; 2]. The overall topology of membrane proteins comprises only the alpha helical proteins and the beta barrel classes, the great majority of membrane proteins belongs to the helical bundle class[3; 4], while those with beta barrel conformation mainly exist as outer membrane proteins (Omps) in Gram-negative bacteria[5]. In my thesis unless otherwise mentioned, MPs mean helical bundle integral membrane proteins.

In addition to their great number, membrane proteins also establish the link across the membrane, implicated in perception and signal transduction, import and export of essential substances. Hence enormous efforts have been dedicated to this protein class. Although in recent years significant progress has been made and the growth of structure determinations of membrane proteins is growing exponentially, in comparison to their soluble counterpart, it is still lagging behind as shown in Figure 1[6].

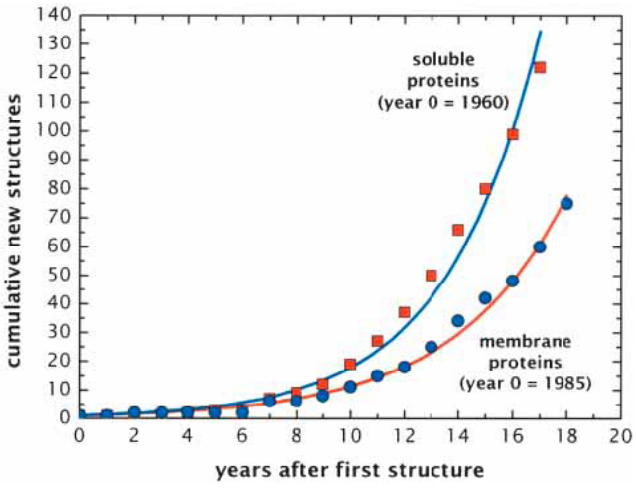


Figure 1. Progress of structural studies of membrane and soluble proteins

1.1 G-Protein Coupled Receptors

G-protein coupled receptors (GPCRs) constitute the largest integral membrane protein family[7]and are one of the largest protein families in the human genome,

accounting for 2% of the total genome[8]. Perhaps more than any other kind of proteins, GPCRs have evolved to recognize a plethora of ligands including ions, organic odorants, amines, peptides, proteins, lipids, nucleotide and even photons can mediate their signal transduction via these proteins. Furthermore GPCRs play critical roles in molecular recognition and signal transduction, which are involved in virtually all biological processes, so members of the GPCR family are amongst the most pursued pharmaceutical targets[9]. Around 30% of all marketed prescription drugs act on GPCRs and around 30% of all targets investigated so far in pharmaceutical companies are focused on GPCRs. This makes this class of proteins the historically most successful therapeutic target family[10]. In spite of the diversity in sequence and function, all members from this superfamily share a common scaffold as shown in Figure 2, which constitutes a 7 transmembrane (TM) helix domain as the core unit (thus GPCRs are also named as 7 TM or heptahelical proteins). 7 TM domains are connected by 3 extracellular loops (ECLs) and 3 intracellular loops (ICLs) of variable length; the N terminus and C terminus are always located extracellularly and intracellularly, respectively. Associated with the receptor is a so-called G-protein complex, which consists of β, γ and a GDP-bound α subunit. Upon activation from the extracellular space, mainly transmembrane helices III and VI of the receptor undergo a conformational changes leading to exchange of the bound GDP by GTP in the catalytic unit of the G-protein. This destabilizes the G-protein, which dissociates from the receptor into activated β/γ and GTP-bound α units, which can have further downstream impact, e.g. on adenylylcyclase or phospholipase C and associated effects on cAMP or Ca^{2+} levels in the cell.

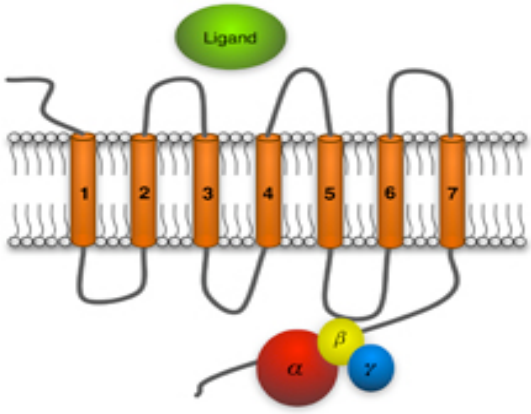


Figure 2. Schematic representation of GPCRs

1.1.1 Classification of GPCRs

The GPCR superfamily is historically classified into three main families: 1, 2 and 3, respectively. Rhodopsin belongs to family 1 and is the most intensively studied GPCR, so family 1 is also called the rhodopsin-like family. The other two main subfamilies are the secretin-like receptor family 2 and the metabotropic glutamate-like receptor family 3[11], as is shown in Figure 3 (refer to www.gpcr.org). The superfamily of GPCRs can also be classified into to five groups based on phylogenetic analysis, named rhodopsin, glutamate, adhesion, taste2 and secretin, respectively[12; 13]. In my thesis the historical nomenclature will be used.

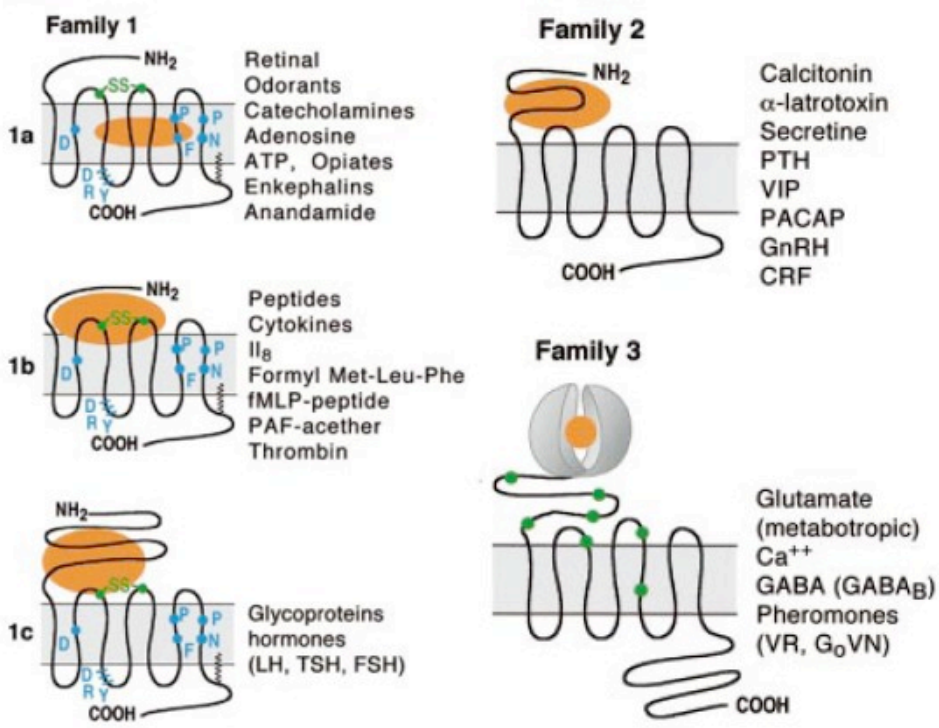


Figure 3. Classification of GPCRs

Family 1 comprises almost 90% of all GPCRs and is by far the largest and best-studied subfamily. Family 1 can be subdivided into receptors for odorants (family 1a), receptors for important neurotransmitters (family 1b), such as dopamine or serotonin as well as neuropeptides, and receptors for glycoprotein hormones (family 1c). Receptors of family 1 are characterized by several highly conserved amino acids (some of them are indicated in the diagram by blue letters) and a disulphide bridge, that connects the first and second extracellular loops (ECLs). Most of these receptors also have a palmitoylated cysteine in the carboxy-terminal tail, which serves as a

membrane anchor. The crystal structure of rhodopsin has indicated that some of the transmembrane domains of family 1 receptors are 'tilted' and 'kinked' due to the presence of certain amino acids such as proline that distort the transmembrane helices.

Family 2 GPCRs have been found in all animal species investigated so far, including mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*, but not in plants, fungi or prokaryotes. Many members of GPCR family 2 contain two additional structural features in addition to the classical 7 TM region: i) a mucin-like region rich in serine and threonine residues; ii) a conserved cysteine-rich proteolysis domain[14]. Their morphology is similar to some family 1 receptors, but the palmitoylation site is missing and the conserved residues and motifs are different from the conserved residues in the family 1 receptors. Little is known about the orientation of the TM domains and their general structure, but given the divergence in amino-acid sequence, they are likely to be different from family 1 receptors. Ligands for family 2 GPCRs include hormones, such as glucagon, secretin and parathyroid hormone.

Family 3 is a relatively smaller subfamily and contains the metabotropic glutamate, the Ca^{2+}-sensing and the GABAB (γ-aminobutyric acid, type B) receptors. A long amino terminus and a long carboxy tail characterize these receptors. The ligand-binding domain is located in the amino terminus and is thought to resemble a Venus fly trap, which can open and close with the agonist bound inside. Except for two cysteines in ECL1 and ECL2 that form a putative disulfide bridge, the family 3 receptors do not have any of the features that characterize family 1 and 2 receptors. A unique characteristic of these receptors is that the third intracellular loop is short and highly conserved. At present, little is known about the orientation of the TM domains and the general structure.

1.1.2 Current State of Structural Studies of GPCRs

There are two major obstacles against high throughput elucidation of membrane protein structures: i) extremely low expression levels in native environment; ii) the chemical inhomogeneity of native membranes and the instability in *in-vitro* membrane mimetics. Concerning the recombinant production of membrane proteins of sufficient quantity enormous efforts have been invested on various expression system such as *E.coli*, yeast, insect cells, mammlian cells and cell free system[15; 16; 17; 18; 19; 20]. Although significant challenges still remain for several GPCRs milligram

quantity can be achieved. Taking advantage of the extensive stacks of specialized membranes and membrane insertion machineries in the photoreceptor cells, Eroglu *et al.* have even tried transgenic flies for expression of the *drosophila melanogaster* metabotropic glutamate receptor, and discovered that amazingly the expression level in such a system is apparently better than in conventionally used systems such as yeast or sf9 cells[21]. So far only 5 out of around 1000 known GPCRs have resulted in high-resolution structures and they come from two subfamilies: bovine rhodopsin (Figure 4a)[22]; bovine opsin (Figure 4b)[23]; squid rhodopsin (Figure 4c)[24]and the human β-2 adrenergic receptor (Figure 4d)[25; 26]or the turkey β-1 adrenergic receptor (Figure 4e)[27]. The available structures confirmed a common architecture of the 7 TM bundle, and the differences among different GPCRs are mainly present in flexible loop regions and the amino or carboxyl terminal domains. Each new GPCR structure revealed the presence of surprising new features. For example, in squid rhodopsin, TM helices V and VI are surprisingly long and deeply penetrate into the cytosol, and a second cytosolic helix is present[24]. The major obstacles to obtaining structures of other GPCRs include protein production and purification, and protein stability and homogeneity. In terms of production, it is now possible to generate sufficient quantities (tens of milligrams) of several GPCRs for crystal screening using bacterial, yeast, insect cell, and mammalian cell expression systems, but for each GPCR a thorough optimization of all parameters are necessary for obtaining milligram yields. The availability of robotic systems for preparing setups of 100 nl volumes (or smaller) has enabled large parameter screens with relatively small amounts of protein. As such, protein production is no longer the major limitation for crystallography efforts. Perhaps a greater problem is the stability of purified GPCRs in detergents compatible with crystallography. For example, the β2 adrenergic receptor (β2AR) and many other GPCRs are not stable in the detergents used to obtain rhodopsin crystals; and rhodopsin crystals have not been obtained in dodecylmaltoside, a detergent in which the β2AR is relatively stable. GPCRs tend to be more stable in non-ionic detergents with relatively long alkyl chains. Another problem is the potential for both structural and conformational heterogeneity in GPCRs. Structural heterogeneity means heterogeneity in posttranslational modifications such as glycosylation, phosphorylation and palmitoylation. These sources of heterogeneity can often be eliminated by site directed mutagenesis of the protein, or enzymatic removal of sugars and phosphates. This source of heterogeneity is minimized if the GPCR can be

expressed in bacteria. The conformational heterogeneity comes from the inherent flexibility of GPCRs, which is presumably important for the function of signal transduction or substance transportation. In addition the loops and the amino and carboxyl termini are intrinsically dynamic due to the lack of interactions with the membrane. The conformational heterogeneity can be mitigated by modification of the sequence or addition of ligands or antibodies as was done for crystallizing the β2 adrenergic receptor.

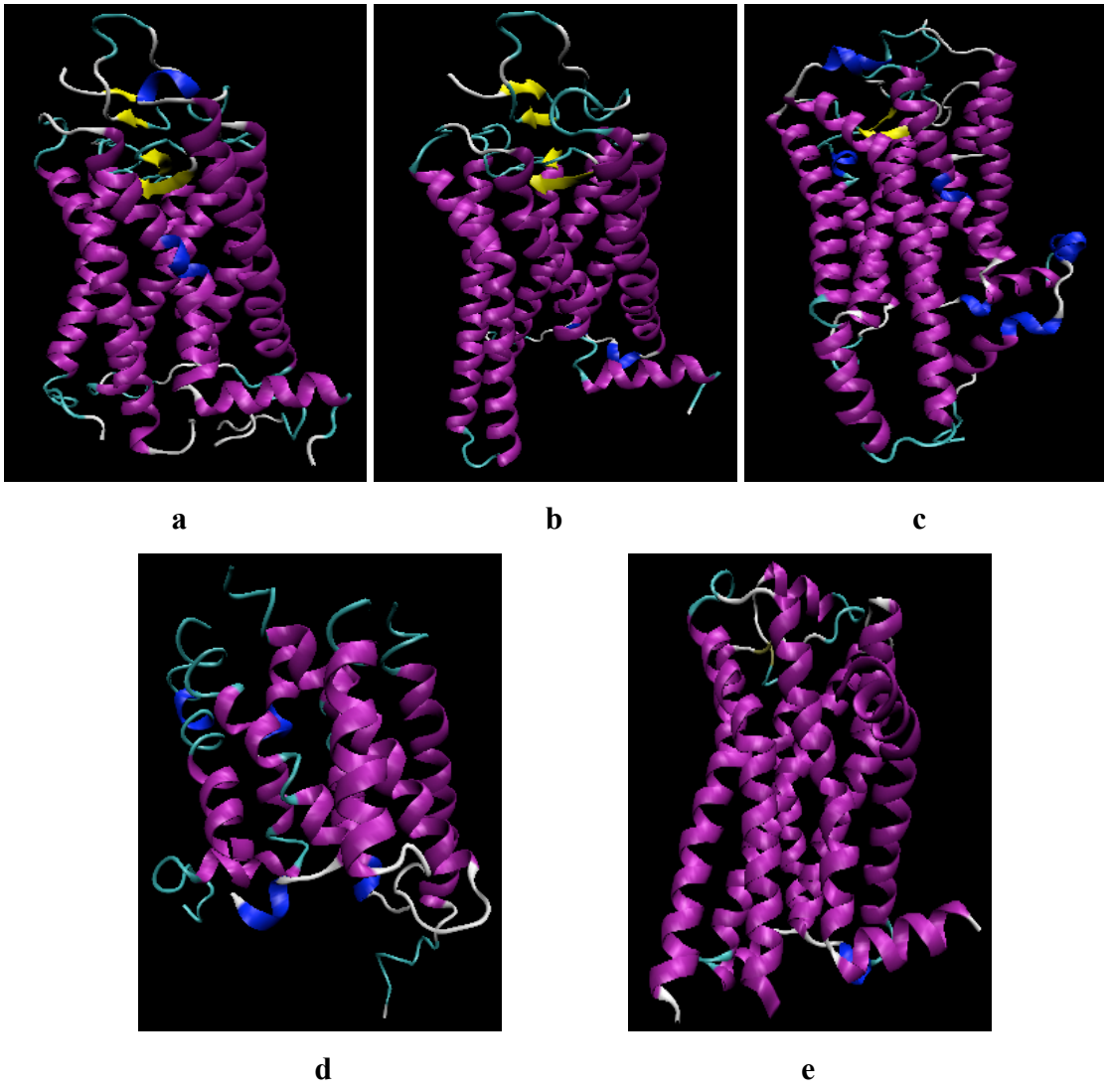


Figure 4. Available high-resolution structures of GPCRs. (a) bovine rhodopsin; (b) bovine opsin; (c) squid rhodopsin; (d) human β-2 adrenergic receptor; (e) turkey β-1 adrenergic receptor

All the high-resolution structures from whole GPCRs were solved by X-ray crystallography, and the application of NMR for membrane proteins requires recombinant production of labeled protein, which imposes limitation on the selection of the expression organism. Nevertheless, the expression in *E. coli* of the human vasopression V2 receptor for subsequent NMR studies has been described[28]. In the

[^{15}N,^{1}H]-TROSY spectrum 230 complete spin systems out of expected 349 were successfully identified, which offers an alternative technique for this fascinating protein family. Quite recently Gautier and coworkers have made another big step in the structural study of a multispan seven-helical membrane protein receptor, sensory rhodopsin pSRII from *Natronomonas pharaonis*, using high-resolution solution NMR methods. In this work near-complete backbone assignment (>98%) of the 241-residue chain has been achieved after careful detergent screening, which finally demonstrated DHPC is the best membrane model[29]. Recently Werner and coworkers have successfully expressed rhodopsin in HEK293 cells in ^{13}C and ^{15}N labeled form[30]. Though this labeling method only covers 49% of all amino acids based on the sequence, it provides a exciting possibility to extend the labeling technique to wide range of organisms, thus promoting the applicability of NMR techniques to more interesting proteins, such like post-translation modified proteins, membrane proteins that can only be expressed in mammalian cells

1.1.3 Oligomerization of GPCRs

The presence and functional significance of GPCR oligomers has been debated for long time, however the pro-evidence keeps accumulating. The concept that GPCRs might dimerize was first provoked by the seminal experiment from Roberto Maggio and Jurgen Wess in which truncated GPCRs (containing transmembrane regions 1–5) from m2 and m3 muscarinic receptor were coexpressed with gene fragments encoding the C-terminal regions of the receptor (containing TM 6–7) and functional receptor activity similar to wt-receptor was recovered when corresponding N- and C-terminus were coexpressed[31] . This approach was further refined when a2/m3 and m3/a2 chimeras were generated in which TM 6 and 7 were exchanged between the a2 adrenergic receptor and m3-muscarinic receptors. Transfection of COS-7 cells with either of the two chimeric constructs alone did not result in any detectable binding activity for the muscarinic ligand N-[^{3}H]methylscopolamine or the adrenergic ligand [^{3}H]rauwolscine. However, cotransfection with alpha 2/m3 and m3/alpha 2 resulted in the appearance of specific binding sites (30-35 fmol/mg of membrane protein) for both radioligands. These sites displayed ligand-binding properties similar to those of the two wild-type receptors, which indicates the dimerization might exit on molecular level[31]. Michel Bouvier, Terry Hebert and coworkers also used epitope-tagged receptors and Western blotting techniques to monitor homodimerisation of the β2-

adrenergic receptor and to identify a peptide sequence corresponding to TM6 that could disrupt receptor dimerization and reduce stimulation of adenylyl cyclase[32]. A clearcut example of a receptor heterodimer that is essential for functional activity comes from the interaction between the $GABA_BR1$ and $GABA_BR2$ gene products [33]. Fiona Marshall and co-workers showed that $GABA_BR1$ is normally poorly expressed on the cell surface, but following expression of $GABA_BR2$ there is a marked expression of heterodimers on the cell surface and ten-fold increase of the agonist potency. Furthermore taking the C-terminal domain of the $GABA_BR1$ receptor at bait against a human brain cDNA library using the yeast two-hybrid system identified the $GABA_BR2$ receptor as the major hit. More recently, fluorescence and bioluminescence resonance energy studies have provided convincing evidence that both homodimers and heterodimers exist in living cells[34]. Perhaps the most visually impressive evidence for dimerisation has come from atomic force microscopy images of rhodopsin homodimers in native retinal discs shown in Figure 5[35].

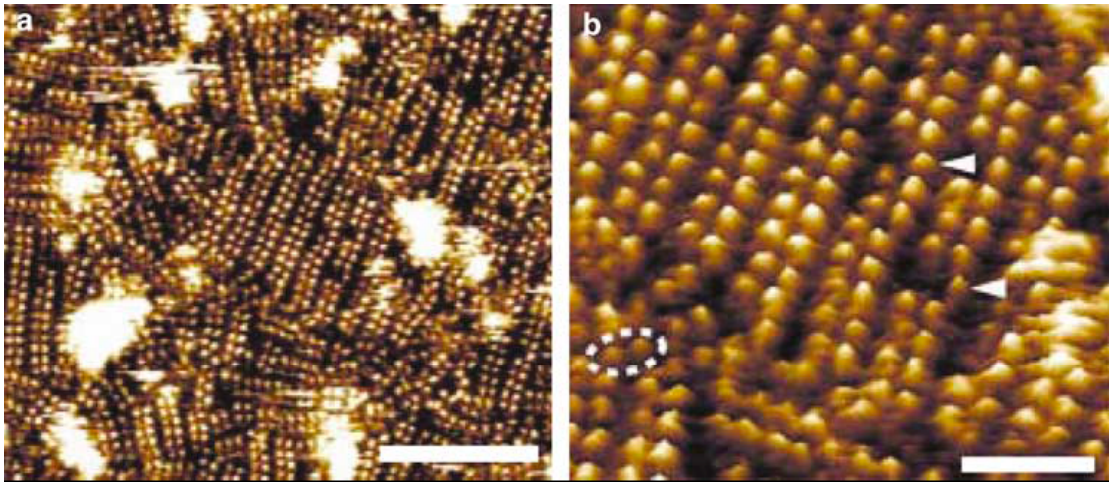


Figure 5. Organisation and topography of the cytoplasmic surface of rhodopsin. (a) Topograph obtained using atomic-force microscopy showing the paracrystalline arrangement of rhodopsin dimers in a native disk membrane. (b) Magnification of a region of the topograph shown in (a) showing rows of rhodopsin dimers. Individual dimers (dashed elipse) and monomers (arrow heads) can also be observed. Scale bars (a) 50nm and (b) 15 nm.

Additionally dimerization of GPCRs can play regulatory roles. George and coworkers have found that μ-opiloid and δ-opioid receptors could be visualized following electrophoresis as monomers, homodimers, homotetramers, and higher molecular mass oligomers[36]. However when μ-opiloid and δ-opioid receptors were coexpressed, the highly selective synthetic agonists for each had reduced potency and altered rank order, whereas endomorphin-1 and Leu-enkephalin had enhanced

affinity, suggesting the formation of a novel binding pocket. In contrast to the individually expressed μ-opiloid and δ-opioid receptors, the coexpressed receptors showed insensitivity to pertussis toxin and continued signal transduction, likely due to interaction with a different subtype of a G protein. Gomez *et al.* have also observed that μ-opiloid and δ-opioid receptor complexes exist in live cells and native membranes and that the occupancy of δ-opioid receptors (by antagonists) is sufficient to enhance μ-opioid receptor binding and signaling activity[37].

1.1.4 Posttranslational Modifications of GPCRs

Posttranslational modification is the chemical modification of target proteins after ribosome translation and is usually the last step for the protein synthesis. Posttranslational modification is a highly effective method found within nature to fine-tune protein activity in a highly sophisticated manner, which of course also adds an additional level of complexity in the topographical organization. Among the available modifications, phosphorylation, glycosylation and palmitoylation have been previously described to play a role in the regulation of GPCR biology.

The phosophorylation of GPCRs are realized by two kinds of kinases: GPCR kinases (GRKs) or second messenger-dependent kinases, such as protein kinase A or protein kinase C. GPCR phosphorylation by second messenger-dependent kinases typically results in diminished receptor-G protein coupling, whereas phosphorylation by GRKs results in both reduced receptor-G protein coupling and enhanced binding of additional proteins (arrestins) [38]. Phosphorylation of GPCRs plays important roles such like desensitization, trafficking, and signaling. Among them phosphorylation-dependent desensitization has drawn enormous attention, since it is an important feature of agonist-stimulated GPCRs that the responsiveness to repeated stimulation with agonists is decreasingly waning. Now the canonical view regarding the desensitization process is as follows: when GPCRs are stimulated with agonists, they activate heterotrimeric guanine nucleotide-binding proteins (G proteins), leading to the production of signaling second messengers, such as adenosine 3',5'-monophosphate, inositol phosphates, and others. Activated receptors are rapidly phosphorylated on serine and threonine residues by G protein-coupled receptor kinases. Phosphorylated receptors bind the multifunctional adaptor proteins beta-arrestin1 and beta-arrestin2 with high affinity. Beta-arrestin binding blocks further G

protein coupling, leading to "desensitization" of G protein-dependent signaling pathways[39]. The phosphorylation sites have been mainly mapped to the carboxy tail.

Glycosylation is another important posttranslational modification of GPCRs where polysaccharides are attached to protein side chains via enzymatic processes. Two kinds of glycosylation are found in nature: N-linked glycosylation, in which the polysaccharides are coupled to the amide group of asparagines and Asn-X-Ser/Thr (X indicates any residue except Proline), is the common motif recognized by the polysaccharides transferases, and O-linked glycosylation in which the hydroxyl groups of serine or threonine are coupled with the polysaccharides. Though N-linked glycosylation is assumed to be an almost universal modification of GPCRs, quite few examples for O-linked glycosylation are available, amongst which are PGE_2 and PGD_2 prostaglandin receptors[40] and octopus opsin[41]. Though a common N-glycosylation motif is present for all members, the influence of such modifications on GPCRs varies from case to case and even in an unpredictable manner. Consequently, it has been reported that oligosaccharide moieties are important for V2 vasopressin receptor expression and/or stability but not for ligand binding[42]. Similarly, glycan chains are essential for correct folding/trafficking of vasoactive intestinal peptide-1 receptor, the thyroid stimulating hormone receptor[43] and follicle stimulating hormone receptor[44]. For some GPCRs including the somatostatin receptor[45], rhodopsin[46], the β2-adrenergic receptor[47] and the gastrin-releasing peptide receptor[48], glycosylation is important for high affinity ligand binding and/or receptor-G-protein coupling. For many GPCRs, however, no functions have been observed related to N-linked glycosylation, and this groups includes the histamine H2 receptor[49], the A_{2a} adenosine receptor[50] and the AT_2 angiotensin receptor[51]. The amino terminus is the fragment where most of the glycosylations have been observed. In fact, in the N-terminal domains from Y receptors there are potential N-linked and/or O-linked glycosylation motifs. The group of Markus Aebi has developed tools to transfer the machinery for N-linked glycosylation encoded in *Campylobacter jejuni* to *Escherichia coli*. The eukaryotic primary consensus sequence for N-glycosylation is N terminally extended to D/E-Y-N-X-S/T (Y, X not equal P) for recognition by the bacterial oligosaccharyltransferase (OST)[52]. In collaboration with the group of Aebi (ETHZ Microbiology) I have also tested *in vitro* N-glycosylation for the N terminus of the Y2 receptor. Unfortunately, the poor yield of the in vitro glycosylation product did not allow us to investigate the effect of glycosylation on structure and function of this

domain.

Palmitic acid is a 16-carbon saturated fatty acid. When it is attached to the sulfhydryl group of cysteine by forming a thioester bond, the process is termed as palmtoylation. In many cases, palmitoylation appears to play an important role in the expression of functional GPCRs on the cell surface. For example, mutation of the palmitoylated cysteines within bovine opsin led to significant intracellular retention of the mutants when expressed in COS cells[53]. Receptor intracellular trapping upon mutating the palmitoylation sites has also been reported for the canine H2 histamine receptor[54]and the CCR5 chemokine receptor[55]. Accelerated degradation has been shown to occur in the palmitoylation mutants of the human A1 adenosine receptor. FLAG-tagged versions of the wild-type and palmitoylation-deficient mutants of this receptor have drastically different half-lives. Approximately 60% of the non-palmitoylated receptor mutant was cleaved into two smaller polypeptides with a half-time of 0.8 hours[56], which strongly suggests that palmitoylated cysteines are important to allow normal processing of GPCRs. Rhodopsin is the first GPCR, in which palmytoylation has been observed and the location is mapped to the cysteine in the carboxyl terminus[57]. Considering the presence of cysteine at similar positions in many GPCRs, it is apparent that cysteine residues proximal to the cytosolic end of the seventh membrane span may represent general targets for palmitoylation. Supporting this assumption it is the fact that the palmytoylation site is located in the cytosolic tail in case of β2 adrenergic receptor[58]. However, please note that GPCR palmitoylation is not limited to the C-termini of the proteins. Indeed, there is an increasing body of evidence that points to the presence of palmitoylated cysteines in the intracellular loops connecting the transmembrane spans. For instance, mutation of all the cysteines in the carboxyl tail of the rat μ-opioid receptor failed to affect palmitate incorporation, indicating that the palmitoylation site(s) resided outside this receptor domain[59].

1.2 Neuropeptide Y Receptors (Y Receptors)

Y receptors are members of the rhodopsin-like GPCR family 1, which are featured by a binding pocket located either within the 7TM bundle (indicated as orange in Figure 1a) or at the extracellular loops; there is a conserved disulfide bond connecting extracellular loop 1 and 2; there are conserved residues in 7TM domains. Y receptors exert their functions via interaction with neurohormone Y (NPY) peptides and the binding is assumed to involve residues from the extracellular loops and possibly the N terminal domain[60].

1.2.1 Evolution and Members of Y Receptors

Y receptors currently encompass five cloned members in mammals, Y1, Y2, Y4, Y5 and y6[61] and they have different binding affinity against different ligands shown in Table 1. The y6 receptor is designated in a lower case due to the fact that y6 is a truncated version of Y receptors in mammals and no physiological function has been associated with it so far[62]. In my research I only focused on Y1, Y2, Y4 and Y5 receptors. The genes for Y1, Y2 and Y5 are clustered together on homo-sapiens chromosome 4 (Hsa4), the Y4 gene is located on Hsa10 and the y6 gene is on Hsa5. These three chromosomes share members of numerous other gene families, supporting the idea that they all arose from a common ancestral chromosome through duplications that took place in an early ancestor[63]. The phylogenetic analyses show that Y1, Y2 and Y5 subfamilies are very distantly related, thus the ancestral chromosome carried a representative for each of these three subfamilies before the chromosome duplications. After the duplications, some genes were lost, but interestingly the gene losses seem to differ between the vertebrate lineages. For instance, mammals have lost Y7 and teleost fishes seem to have lost Y1, Y5 and Y6[64]. Even though all the Y receptors evolve from the same ancestor, they differ from each other more than any other G protein coupled receptors with only 27-31% overall identity (shown in Figure 6). It is even more surprising in consideration of the fact that Y1, Y2 and Y5 receptors bind to not only one but two endogenous ligands, Neuropeptide Y (NPY) and peptide YY (PYY)[65]. While the Y4 subtype is clearly more similar to Y1 with 43% overall identity, Y4 is considered as the pancreatic

polypeptide (PP) preferring receptor[66] even though Y5 also has slightly lower affinity to PP[67]

Table 1. Inhibition constant (nM) of neuropeptide at Y receptors

	Y1	Y2	Y4	Y5	Type[a]	Reference
NPY	0.81	0.02	1.90	0.19	K_i	68
PP	> 1000.00	> 2000.00	0.04	58.00	IC_{50}	67
PYY	1.10	0.01	1.06	0.62	K_i	68

[a] Type of inhibitory constant quoted. Ki=inhibition constant, IC50=concentration at which 50% inhibition occurs.

1.2.2 Physiological Functions of Y Receptors

The Y1 receptor is the first neurohormone binding receptor to be cloned and widely expressed in the central nervous system including the major nuclei in the hypothalamus. It is thought to be responsible for the regulation of food intake and energy homeostasis. Intracerebroventricular injection of the Y1/Y5 preferring ligand Leu31/Pro34 NPY into rodents strongly stimulates feeding behavior. Interestingly, Y1 receptor knockout models do not display any major abnormalities in regard to food intake or bodyweight, although there are subtle changes observed in all Y1 receptor knockout mice lines analyzed. Many physiological functions of NPY are mediated via the Y1 receptor, for example the vasocontraction function of NPY is completely abolished in Y1 knockout mice[69]. Recently, the Y1 receptor was also observed to be involved in the regulation of voluntary alcohol consumption[70], and it plays an important role in autoimmune and inflammatory processes[71].

The Y2 receptor is widely expressed in the central nervous system (CNS) with particularly high levels found in an area of the arcuate nucleus with a permeable blood brain barrier. This makes the Y2 receptor accessible to circulating factors and an ideal candidate for mediating peripheral signals on the regulation of energy homeostasis. The assumption has recently been confirmed by the discovery of the endogenous Y2 agonist PYY(3-36) that is released from the gastrointestinal tract postprandially[72]. Similar to the Y1 receptor the Y2 receptor is also directly related to some functions from NPY[73]. For instance, in pig spleen, a Y2-specific agonist evoked potent vasoconstriction that could be inhibited by a Y2-selective antagonist, BIIE0246. In addition, the Y2 is involved in NPY-induced angiogenesis and circadian

rhythms[74]. Knockout studies of the Y2 receptor have shown that this receptor may be involved in the feeding response to NPY as well as in bone formation[75].

The Y4 receptor, also known as the PP preferring receptor, is mainly expressed in peripheral tissues including colon, intestine, prostate and pancreas. Unlike other Y receptors, the Y4 receptor demonstrates poor sequence conservation, which means that the Y4 is one of the fastest evolving GPCRs known. The Y4 knockout mice show reduced food intake and significantly reduced body weight[76]. Plasma levels for PP are strongly elevated and white adipose tissue mass is reduced in male knockout mice. The increases in plasma PP levels in these animals are the most likely explanation for the reduced food intake and body weight seen in these mice. Centrally located Y4 receptors may be involved in the regulation of reproduction as the Y1 antagonist/Y4 agonist 1229U91 induced release of the luteinizing hormone when injected into the spinal cord[77].

The Y5 receptor is expressed in the human hypothalamus, with the highest density being found in the arcuate nucleus. With the administration of an antagonist, the Y5 receptor has been proved to be involved in the regulation of NPY-mediated food intake[78], however similar to Y1 receptor, the knockout mice can eat and grow normally, which does not confirm this point[79].Furthermore, it has been shown that the Y5 selective agonist [D-Trp32]NPY inhibited neuronal activity in the suprachiasmatic nucleus without generating a phase-shift indicating that Y5 may also be indirectly involved in regulation of circadian rhythms[77].

In addition to the regulation of food uptake, recently Y receptors have been found linked to neoplasia. Y receptors were also found to be over-expressed in a variety of human cancers, mainly expressed in specific endocrine tumors and epithelial malignancies as well as in embryonal tumors[80; 81; 82]. Preliminary experimental data suggests that tumoral NPY receptors may be functional *in vivo*. They may be activated by intratumoral NPY peptide and may mediate NPY effects on tumor growth and tumoral blood supply. Moreover, NPY receptor expressing tumors are promising candidates for an *in vivo* NPY receptor targeting with radiolabeled and cytotoxic NPY analogs, analogous to somatostatin receptor targeting[83; 84].

Multiple sequence alignment of human Y receptors sequences

Figure 6. Human Y Receptors Multiple Sequence Alignment

1.2.3 Mutagenesis Studies of Y Receptors

Based on pharmacological studies, Y receptors have been proven to be important members in the food consumption network, so they become preferable targets for treatment of relevant diseases, obesity in particular[85]. Many specific agonist and antagonist have been synthesized and demonstrate reasonable efficacy, unfortunately none of them has passed clinical trials till 2006 yet[86; 87]. Furthermore, many studies have been done in exploration of the motifs within Y receptors, which are involved in ligand binding and selection. Mutagenesis studies have been extensively applied to the human Y1 receptor, and many residues important for ligand binding have been identified (shown in Figure 7). Sautel and coworkers demonstrated that the hydrophobic binding pocket of human Y1 formed by F41 and L43 in TM1, F96 and Y100 in TM2, F286 in TM6, H298 in TM7 is critical for NPY binding, especially the Y100F Y1 mutant lost the binding ability completely, the corresponding part in NPY is the hydrophobic residue Y36 located in the C terminus[88]. Walker and co-workers found that acidic residues located in the extracellular loops of Y1 receptor are important elements for interactions with positively charged residues in NPY, in particular those located in the C terminus[89]. In mouse two *in vivo*-expressed splice variants of Y1 receptor have been found [90]. The short form (307 amino acids) of the Y1 receptor ends a few amino acids after the third extracellular loop, yielding a receptor with an incomplete TM7. NPY binds to this short form of the Y1 receptor with similar affinity as to the complete 384 amino acid protein. However, the signaling of the short form of the receptor was impaired, indicating that the TM7 and the carboxyl-terminal tail are not essential for ligand interaction but rather for G-protein activation. When positions 31 and 34 of NPY or PYY (Ile and Gln) are replaced by the corresponding amino acids in PP, Leu, and Pro, respectively, the resulting peptides do not bind to Y2 receptor anymore, although this peptide remains a potent full agonist at the other PP-fold receptors[91]. It was later shown that only the Pro34 substitution was essential for preventing Y2 receptor binding[92]. NPY and NPY2-36 are equally potent against Y5 receptor in producing a large increase in feeding after intracerebroventricular administration, suggesting a different mechanism other than Y1, Y2 and Y4. In addition, NPY with position 32 replaced with D-tryptophan ([D-Trp32]NPY) selectively inhibited NPY-induced feeding, and it was found that [D-Trp32]NPY is a modestly selective agonist at Y5 expressed in HEK293 cells acting to inhibit cAMP synthesis but with a lower potency than NPY, PYY and

NPY2-36[93]. Similarly Beck-Sickinger has mutated all the conserved residues in the Y5 receptor across different species with alanine and identified those that affect the binding with neurohormones significantly. Further more taking use of complementary mutagenesis between Y5 receptor and NPY, two interaction partners have been pinpointed: R25(NPY) and D2.68(Y5); R33(NPY) and D6.59(Y5)[94].

Some of the differences between the different mutagenesis studies can probably be explained by the very heterogeneous expression systems used. For instance, Walker and colleagues used a vaccinia virus vector to express the human Y1 receptor in Hela cells[89]and later found differences when these mutants were compared with the same mutants expressed in E. coli and mammalian cell lines[95]. It is also possible that the usage of different mammalian cell lines may alter the pharmacology of the very same receptor protein depending on what other proteins are expressed in the cells. In summary the interaction between NPY hormones and Y receptors is complex and involves hydrophobic interactions and electrostatic interactions, and it possibly involves the presence of other participants such as the membrane.

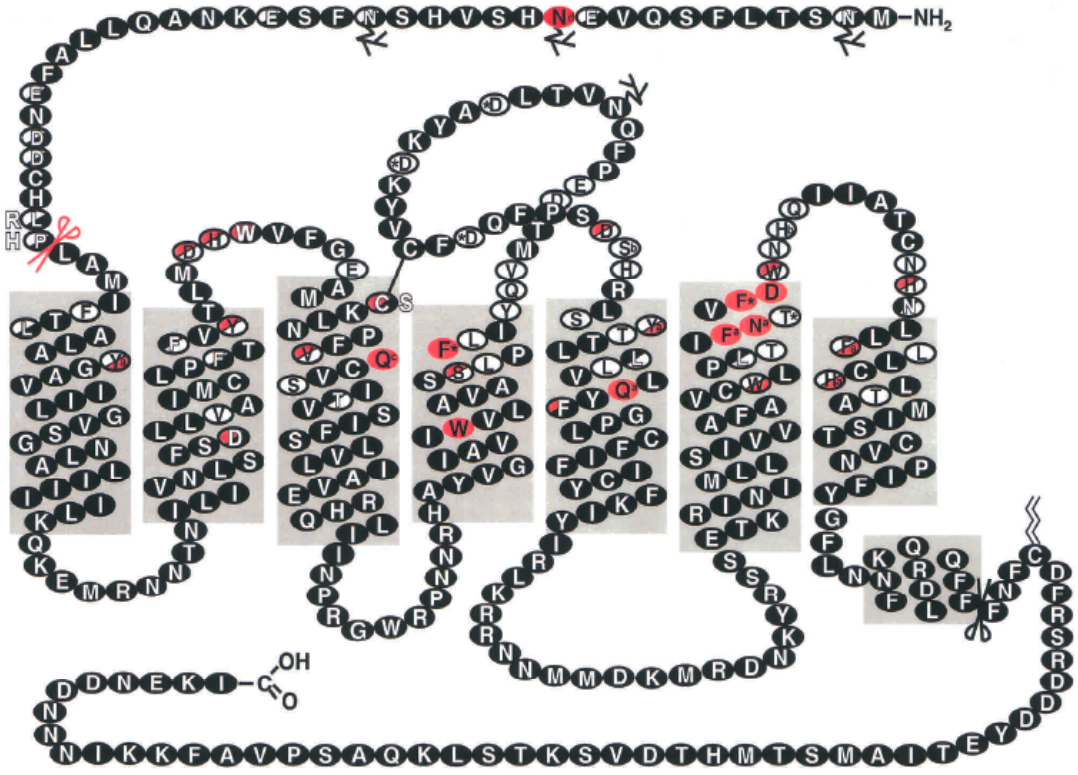


Figure 7 Summary of structural and mutagenesis studies performed on the mammalian Y1 receptor. Amino acids were mutated to alanine except in three positions. Those are indicated by a shaded amino acid next to the original sequence. Red indicates a loss in affinity >2.5-fold. Upper left: effects on NPY/PYY binding, lower right: antagonist. White: no effect. Black: not done.

1.2.4 Which Portions of the Y Receptors are Important for Ligand Binding?

Most of the studies have emphasized the importance of extracellular loops, transmembrane domains, intracellular loops and C-terminal domains in ligand-binding and signal transduction. However, little is known about the role of the N terminal domain in GPCRs. In GPCR family 2 and 3, the N terminus is known to be involved in the ligand-binding process (shown in Figure1b and 1c), the N-terminal domain of GPCR family 2 shares common features: typically 6 conserved cysteines, 2 conserved tryptophans and an aspartate acid which has been suggested to be critical for ligand binding[96; 97]. Sun and his coworkers have solved the solution structure of the N-terminal domain from pituitary adenylate cyclase-activating polypeptide (PACAP) receptor complexed with its ligand PACAP[98]. But in family 1, the rhodopsin-like GPCRs, little attention has been devoted to that issue.

The N terminal sequences among the four Y receptors are not conserved (shown in Figure 8), however relatively high sequence similarity can be observed between Y1 and Y4 (shown in Figure 3), which is consistent with a phylogenetic study. The non-conserved sequences in the N termini among the whole Y receptor family indicate that they may play various roles in the corresponding receptors.

```
N-Y1        MN-STLFSQVENHSVHS-NFSEKNAQLLAFENDDCHLPLAMI
N-Y4        MNTSHLLALLLPKSPQGENRSKPLGTPYNFS-EHCQDSVDVM
            ** * *:: :  :* :. * *:  .    *. :.*: .: ::
```

Figure 8. Sequence Alignment between N-Y1 and N-Y4

Based on a secondary structure prediction, the N terminal domains display dramatic disorder as shown in the Figure 9. The bioinformatics prediction does not necessarily imply the lack of structure, especially because it does not take the presence of a membrane into account. Furthermore, disorder often occurs in nature, and exits in many proteins[99]. Disordered regions, according to statistics, are also prominent in ligand recognition[100]. The lack of folded structure in signaling proteins might give these proteins a functional advantage over globular proteins with well-defined secondary and tertiary structure: the ability to bind to multiple different targets without sacrificing specificity and the ability to overcome steric restrictions, thus enabling larger surface interactions[101].

```
N-Y1:  MNSTLFSQVENHSVHSNFSEKNAQLLAFENDDCHLPLAMI
       ------EEH----------HHHHHHHH-------------
N-Y2:  MGPIGAEADENQTVEEMKVEQYGPQTTPRGELVPDPEPELIDSTKLIEVQ
       -------------HHHHHH--------------------H----HEE--
N-Y4:  MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVM
       ----HHHHH--------------------------------E
N-Y5:  MSFYSKQDYNMDLELDEYYNKTLATENNTAATRNSDFPVWDDYKSSVDDLQ
       --------------HHHHH------HH-----------------------
```

Figure 9. Secondary Structure Prediction Based on Primary Sequence as predicted by the program nnPredict (http://alexander.compbio.ucsf.edu/~nomi/nnpredict.html) H = helix, E = strand, - = no prediction

In the pioneering study, Monteclaro fused the N terminal domain from CCR2 (member of the GPCR family 1) receptor to the human CD8 transmembrane and cytoplasmic domain. The chimera protein was expressed on the HEK-297T cell surfaces and binding affinity towards its ligand, the monocyte chemoattractant protein 1 (MCP1) was measured by ELISA. The results surprisingly suggested that the N-terminal domain alone is both necessary and sufficient for ligand binding. Based on the results, a two-step activation model was proposed, in which the noncovalent "tethering" of MCP-1 by the receptor amino terminus enhances the low affinity interactions with extracellular loops and eventually leads to G-protein activation [60]. Based on a mutagenesis study, Ho and co-workers concluded that the N terminus of Kaposi's sarcoma-associated herpesvirus GPCR, also a member of GPCR family 1, is necessary for high affinity chemokine binding but not for constitutive activity[102]. In Prado's work, the N termini from chemokine interleukin 8 (IL-8) receptor CXCR1 and CXCR2 were demonstrated to play a role in the interaction with and internalization of their ligands[103]. Andersson, in a study about the membrane assembly of the cannabnoid receptor 1 (CB1, member of GPCR family 1), also proposed that the N terminus might play a role in regulating the stability and surface expression of CB1[104]. In Dong's work, the mutation of Lys 187 within the second extracellular loop led to silencing of signal tranduction instead of ligand binding, in addition the truncation of the N terminus can eliminate the negative effect of the mutation, which indicates there is a acidic residue in the N terminal domain that interacts with the basic residue lysine during signal transduction[105]. Gupta and coworkers recently found the antibodies raised with the N-terminal domain from μ-opioid receptor (GPCR 1) can recognize the activated receptor with higher efficacy, indicating the N-terminal domain adopts a conformational change upon ligand binding, and it might also be involved in the direct ligand binding[106]. Regarding the

structure of the N-terminal domains, the crystal structure of rhodopsin reveals five distorted strands among which the first two antiparallel strands form a typical β sheet conformation[22].

Since our preliminary model indicates that the N-terminal domain might be a functional unit from Y receptor that can be involved in ligand recognition or binding. Besides, many other groups have also found the importance of the N-terminal domains from various GPCR families including family 1. Therefore in my Ph.D. studies, I started from all the N terminal domains of Y receptors.

1.3 The Ligands of the Y Receptors: Neuropeptide Y Hormones

Neuropeptide Y (NPY) hormones are members of orexigenic neuropeptides and they are distributed in the central nervous system as well as in the peripheral nervous system. Since they play important roles in many physiological processes, such as regulation of food uptake, circadian rhythms and endocrine levels, a lot efforts have been devoted to these.

1.3.1 Members of NPY Family

NPY together with peptide YY (PYY) and pancreatic polypeptide (PP) form the NPY family pepitdes, which are characterized by 36 amino acid residues and amidation at the C terminus[107]. NPY is well known to be the most abundant neuropeptide in the central nervous system in mammals, although it is also widely expressed in the peripheral nervous system[108]. NPY has been functionally implicated in feeding behavior, cardiovascular regulation, control of neuroendocrine axes, affective disorders, seizures, and memory retention, tumor progression[109], with food intake regulation the most noticeable[110].

PYY, sharing high sequence identity to NPY, is produced by the intestinal L-cells, the highest tissue concentrations of PYY are found in distal segments of the gastrointestinaltract, although it is present throughout the gut[111]. Circulating PYY exists in two major forms: PYY1–36 and PYY3–36. PYY3–36, the peripherally active anorectic signal, is created by cleavage of the N terminal Tyr-Pro residues by dipeptidyl peptidase 4 (DP4)[112]. PYY3–36 is a major form of PYY in both the gut mucosal endocrine cells and the circulation, however I will focus only on the full length PYY in my study. PYY is released into circulation following food intake, the concentrations are proportional to meal energy content, so that higher levels are seen after fat-intake as compared to sugars and proteins[113].

PP is predominantly located in endocrinic cells in the pancreas (the reason why it is named pancreatic polypeptide) and released into circulation as part of its preprotein[114]. Similar to NPY and PYY, it is involved in gastric emptying, glucose metabolism and insulin secretion[115]. In my research bovine PP (bPP) is used instead of human PP because it has been structurally characterized by Dobson[116] and they

differ only in position 6 (V from human to E from bovine) and 23 (D from human to E from bovine).

In addition to the neuropeptides mentioned above, there exist a truncated version of NPY and PYY, which in contrast to the full-length peptide start from position 2 or 3 and more importantly loose their efficacy at Y1 receptor while they are active especially at the Y2 receptor. Dipeptidyl protease 4 (DP4) might be the major cleaving enzyme[117]. NPY3–36 and PYY3–36 have been shown to play a role in energy metabolism via inhibition of exocrine pancreas function[118]or other feeding associated processes[119], and probably are involved in several other as yet to be discovered physiological functions. These regulatory processes are closely dependent on the expression and function of DP4-like peptidases due to their capability in hydrolyzing the post-proline bond between positions 2 and 3 of NPY and PYY.

1.3.2 Structural Studies of NPY Members

The first crystal structure of NPY family member, avian pancreatic polypeptide (aPP), was solved in 1981 by Blundell (shown in Figure 10) and was also the first high resolution structure of smaller peptide[120]. In the structure, the N terminus adopts a polyproline type II helix comprising residues 1-8, the C terminal residues 14-31 form the typical α helix, the N terminal part is connected by a β turn and bent back onto the C terminal part. This kind of structural motif is since then named as “PP fold” and confirmed by the solution structure of bPP elucidated by NMR[116]:

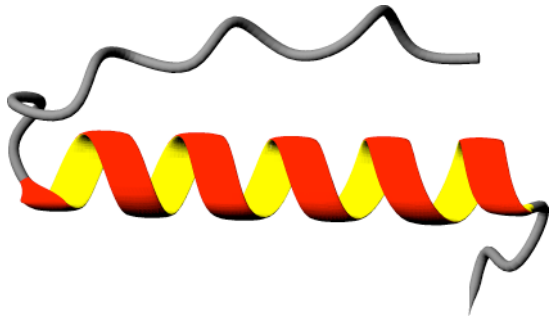


Figure 10. Crystal structure of avian pancreatic peptide (1PPT)

All three neuropeptides share high sequence identity, especially between NPY and PYY 69% sequence identity is observed as shown in Figure 11, and the sequence homology is larger than 80%. Considering the relative high sequence identity, it is expected that the PP fold is a common and characteristic structural motif amongst

peptides of the NPY family, however later studies demonstrated surprising conformational diversity.

```
NPY      YPSKPDNPGEDAPAEDLARYYSALRHYINLITRQRY-NH2 36
PYY      YPAKPEAPGEDASPEELSRYYASLRHYLNLVTRQRY-NH2 36
bPP      APLEPEYPGDNATPEQMAQYAAELRRYINMLTRPRY-NH2 36
          * :*: **:::*..*:::::* : **:*:*::** **
```

Figure 11. Multiple sequence alignment among neuropeptides

However, peptides with highest sequence identity to PYY, hNPY[121] and pNPY[122] revealed a unstructured N terminus in solution. In consideration of the fact that NPY and PYY have almost the same Y receptor affinity, it is surprising that NPY and PYY adopt dramatically different conformations in solution as shown in Figure 12. This raises the question as to what is the biological relevance of the PP fold in Y receptor recognition? Are there other elements that contribute to ligand recognition?

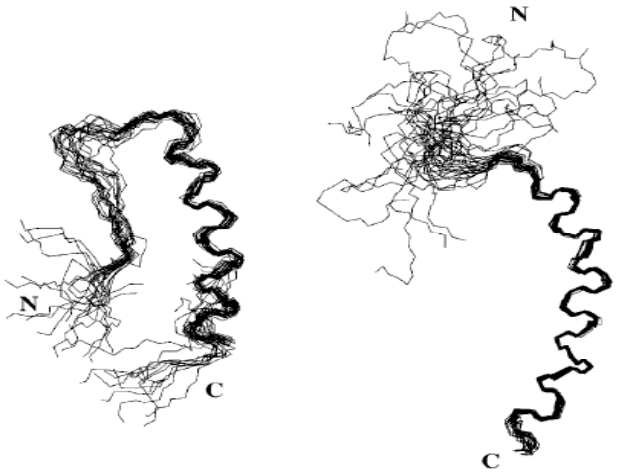


Figure 12. Bundle of the NMR-derived conformers of PYY (left, 1RU5) and NPY (right, 1RON)

In our group, the structures of NPY, PYY, bPP have been determined both free in solution and when bound to dodecylphophocholine (DPC) micelles as shown in Figure 13[123; 124; 125]. DPC is zwitterionic detergent, that has been widely used for mimicking mammalian cell membranes[126]. DPC micelles play very important roles in regulating the structure and dynamics of the neuropeptides. In presence of DPC micelles, a common conformation is present in which the N terminus is diffusing freely in solution indicating the PP fold is disrupted. Using a micelle-integrating spin-label experiment it was shown that the C terminal helix is clearly located parallel to the micelle surface and positioned in the lipid-water interface. According to the

$^{15}N\{^{1}H\}$-NOE data that can be used to determine the backbone stability, it was confirmed that DPC micelles destabilize the N-terminal part of all three neuropeptides. Interestingly, the C termini of NPY and PYY become more rigid after binding to the DPC micelles, hence extending the helix to comprise the whole C terminus, while binding to the DPC micelles does no change the fold of the C terminus of bPP significantly. The C terminal pentapeptide is believed to be important for receptor activation, presumably through electrostatic interaction[88]. R33 and/or R35 in NPY are believed to be the key residues involved in direct contact with acidic residues in Y5 receptor, in another word impose specificity toward Y5 receptor[127]. Beck-Sickinger's work is consistent with this point which suggests R33 and R35 are critical for NPY to bind to Y1 receptor while R35 and Y36 are important for binding to Y2 receptor, hence indicating different binding conformations adopted by NPY at the Y1 and Y2 receptors[128].The similarity between the conformations of NPY and PYY in presence of DPC accounts for the similarity in affinities at the various Y receptor subtypes, and hence may indicate a common mechanism for receptor recognition. In addition it is consistent with the membrane compartment model[129; 130] that the DPC-bound conformation is important for Y receptor recognition.

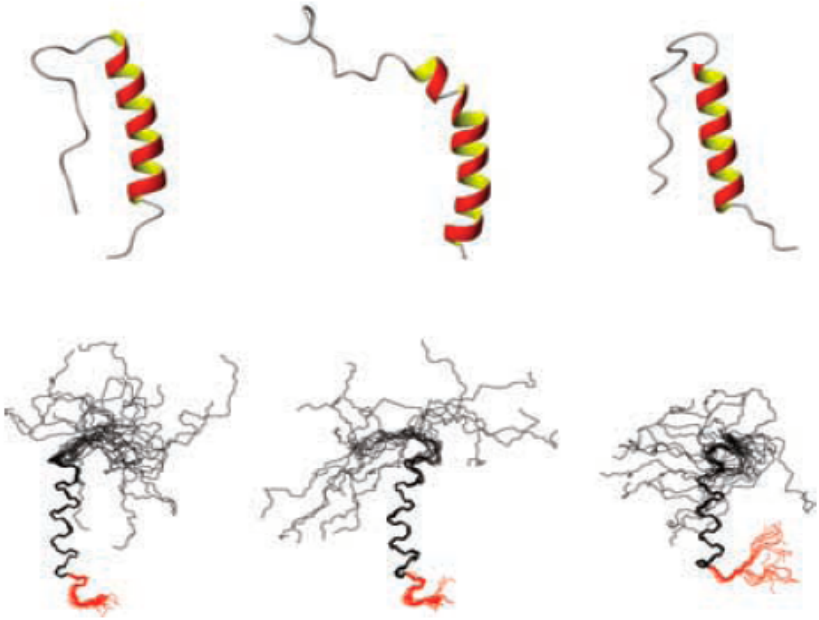


Figure 13. Comparison of structures of PYY (left), NPY (middle), and bPP (right) free in solution (top, single conformer) and when bound to DPC micelles (bottom, superposition of the NMR ensemble). The C-terminal pentapeptide of the micelles-bound peptides is depicted in red.

1.3.3 A Novel Binding Model of NPY Hormones with Y Receptors

In 1984 Kaiser and Kezdy had proposed a pioneering idea that ligands that bind to membrane-embedded receptors recognize their targets from the membrane-bound state[131]. In their research they realized that most of the binding pocket in receptors can only accommodate less than five residues and that the binding is mediated by stereospecific interactions between the ligands and the receptors. However, most peptide hormones and toxins comprise around 30 residues and many of them are not structured in solution but form amphiphilic helices in a membrane environment[132]. Finally they concluded that the amphiphilic domain were responsible for docking the ligands on the membrane where the ligands were recognized by their ligands. Schwyzer et al. developed the concept further into the membrane compartment model[129; 130]. He proposed that ligands were recognized by their receptors after binding to the membrane. Upon binding to the membrane, the ligands were located in proximity to the receptors thereby increasing their local concentration and the receptor search from 3-D to 2-D. Thereby the interaction probability between ligands and receptors is increased. Moreover, conformations close to the bioactive form were induced and the entropic loss upon receptors binding was reduced.

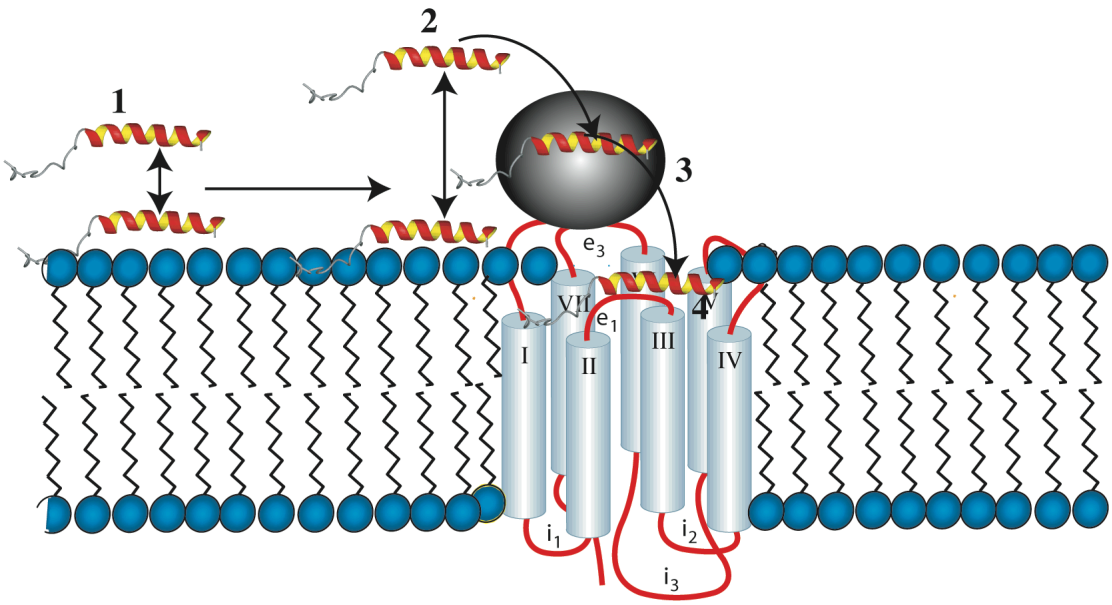


Figure 14. Novel ligand binding model for peptides of the NPY family to their Y receptors

Based on the knowledge described above, a binding model for neuropeptides was proposed in our group as shown in Figure 14[133]. Firstly, the hormones are attracted by the membrane through electrostatic interactions. The extent of these attractions is regulated by the content of the negatively charged phospholipids in the membrane composition and by the content of cationic ligand residues. In a second step, the peptide reorients such that hydrophobic residues penetrate the hydrophobic interior. The hormones subsequently diffuse laterally along the membrane and it is this particular state from which the peptide is initially recognized by the receptor. Our SPR data revealed a moderate binding affinity of the hormones to the membrane, which indicates that the membrane-bound hormones can dissociate from the membrane and diffuse into the binding pocket while still keeping enough affinity to the membrane to guide them the receptor[134]. In the last step, the hormone that is in equilibrium of a membrane-bound and a membrane-water interface state can diffuse into the binding pocket of the receptors. A big difference between membrane-water interface and water phase is observed, e.g. for the dielectric constant. Out data (unpublished) also present a preserved membrane-bound structure in the water/methanol mixture even when the dielectric constant is considerably increased.

1.4 Using Protein Fragments to Study Membrane Protein Structure

The folding studies of soluble proteins have been the focus of many labs, books and conferences. However the progress towards an understanding of the underlying chemical and physical factors is still slow. Membrane proteins are featured with the presence of a stretch of apolar residues and the genome sequencing projects have revealed that membrane proteins comprise 20-30% of the total genome in various organisms and that they play a wide spectrum of functions critical for various physiological processes. At present, people studying folding might think that membrane protein folding could be easier to understand compared to folding of soluble proteins because the membrane environment has significantly restrained the conformational space thus simplifying the complete process and the subsequent analysis.

1.4.1 The 2-Step Membrane Protein Folding Model

In 1990 Popot has proposed a 2-step model to describe the folding of membrane proteins *in vivo*[135; 136]: first partitioning of unfolded polypeptides into the water-membrane interface results in increased formation of backbone hydrogen bonds resulting in formation of secondary structure. Interactions of the side chains with the lipid environment lead to the insertion of transmembrane helices into the bilayer interior. Specific helix-helix interactions lead to formation of native tertiary structure (shown in Figure 15). In this model the TM domain insertion process is separated from tertiary and quaternary structure assembly. Accordingly elements of secondary structure can be thought of as independent folding units/or even functional units and hence be studied separately. Even though the model is simple, a large body of literature has demonstrated supporting information for it: the fragments of various membrane proteins stemmed from proteolysis[137], chemical synthesis[138], recombinant expression through plasmids[138; 139; 140; 141] and cRNA[142; 143] have proved to be capable of reconstituting into functional moieties in lipid vesicles or in various in vivo systems. In addition many structures of the membrane protein fragments have proved to be similar to corresponding parts in the whole protein.

As mentioned above the majority of MPs adopt alpha helical conformations, hence the transmembrane helix is the predominant secondary structure and the helix-loop-helix, the so-called helical hairpin is then the most important tertiary structure as well as a basic folding unit.

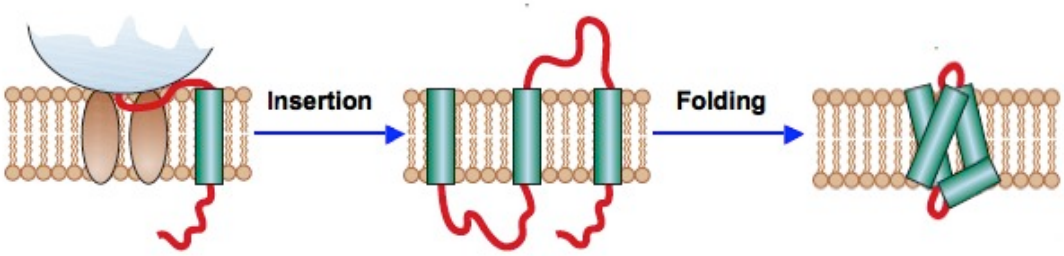


Figure 15. 2-step folding model for membrane protein

1.4.2 Current Progress in Studies of Transmembrane Domains

GPCRs are characterized with the 7 hydrophobic transmembrane helices, therefore many programs have been developed to predict the transmembrane borders on basis of hydrophobicity, which is sufficiently accurate in many cases. For example, a program called TM finder[144] makes use of both hydrophobicity and helicity scales and can identify transmembrane domain with 98% accuray from the primary sequence when applied to the sequence database of membrane proteins. Because there are still ambiguities from the choice of different thresholds and the window size[145],other parameters in addition to the hydrophobicity scale should also be employed to provide further support for identification of the TM borders.

Based on artificial peptide it has been determined that the minimal transmembrane domain of a helical hairpin should comprise 18 residues[146]. Transmembrane domains encompassing residues between 18 and 36 have been observed[147], the average transmembrane length is 20.3 residues[148]. The charged residues in short loops connecting the hydrophobic segments are potent topogenic determinants. The best known rule is the “positive-inside rule”, that states that loops containing positive residues tend to remain at the cytoplasmic side. This rule can not only predict the borders of a transmembrane helix[149] but may also determine the orientation of transmembrane domains in the lipid environment[150]. In consideration of the considerable differences in charges and hydrocarbon length of the detergents we are

using, the orientation of the TM is still unpredictable. Aromatic residues have been observed to tend to cluster at or close to the water membrane interface[151], and such orientational preferences can serve as the driving force for folding and stability of the transmembrane domain[152]. On basis of site-directed and saturation mutagenesis, Yuan has suggested that Tyr can be a primary recognition element for precise transmembrane positioning[153]. Although aromatic residues can be a strong anchor at the water membrane interface, Trp and Tyr but not Phe can also pull the transmembrane helix out of the membrane when positioned in the center[154]. Proline is undoubtedly important for the function and structure of membrane proteins as evidenced by the observation that Pro has one of the highest genotypic propensity in the analysis of transmembrane sequence from the Human Gene Mutation Database[155]. Due to the lack of the amide proton, proline is well known to be the helix breaker in soluble proteins, it conducts similar role in membrane proteins as well[156]. According to statistics proline occurs frequently in the center of a transmembrane helix, where it induces the formation of molecular hinges or kinks, which usually locate about 4 residues N-terminal of the proline residue[157]. Though proline is not necessarily required for kink helices, the sites where proline resides is even slightly conserved in the aligned sequences of transmembrane domains are likely to be in kinked conformations across all the related proteins[158].

Now there are more and more studies focusing on helical hairpins. Hessa *et al* by using *in vitro* translation of a systematically designed model protein in presence of dog pancreas microsomes, have presented a biological hydrophobicity scale, which displays a nice correlation with the biophysical hydrophobicity scale, they also determined the position-dependent contribution of the 20 naturally occurring amino acids to facilitate insertion of the helix into the membrane[159; 160]. Johnson *et al* have proven that van der Waals packing is playing a crucial role for helical hairpin insertion and stability by designing a small protein which complies with the "knobs into holes" rule[161]. In addition aromatic and cation-π interactions can also enhance helix-helix interaction in a membrane environment[162]. Due to the much lower dielectric constant in the membrane environment, hydrogen bonds become significantly stabilizing for folding, dynamics and helix-helix interactions[163]. Furthermore, mutation in the transmembrane domain into polar residues can lead to disease because of the additional hydrogen bond interaction, such like a valine to aspartate mutation in TM4 in the cystic fibrosis transmembrane conductance regulator

protein[164]. Taking a poly-Leu peptide as a model protein, Magnus and his coworkers have determined the minimal length of a helical hairpin, which is 31 amino acids[146] and the turn forming propensity scale in membrane environment for each of the 20 amino acids, which is shown to be largely different from the turn forming propensity scale for globular soluble protein[165]. Akos stressed the significance of turn propensity by suggesting that in case of a naturally occurring helical hairpin from the secretary Na^+-K^+-$2Cl^-$ cotransporter NKCCl the formation of the hairpin structure is only dictated by a short, but strong signal for turn formation instead of specific helix-helix interactions[166]. By comparison of the turn region of cystic fibrosis transmembrane conductance regulator between mutants and wide-type, Hania *et al* have observed a correlation between decreased migration rate in gel electrophoresis and increasing helix content, which indicates that the turn region influences the structure of membrane proteins[167]. Marika further mentioned that charged residues, especially lysine and aspartate in the loops regions can promote the formation of the transmembrane helix[168].

Because of the wide spectrum of difficulties involved in the crystallization of membrane proteins, nuclear magnetic resonance (NMR) becomes a valuable alternative technique. The less stringent requirements on the sample have allowed NMR to be widely applied to structural studies on MPs [28; 169; 170; 171; 172; 173; 174]. More importantly the NMR-determined structures of the isolated loops and transmembrane domains displayed a surprising agreement with the already known crystal structures of the whole proteins[175; 176; 177], which strongly supports the fragment-based approach. With the advance of NMR spectroscopic techniques such as TROSY, the generally considered molecular weight limitation has been dramatically lifted, and the so far largest protein studied by NMR is a bacterial chaperon GroEL-GroES complex, which has molecular weight of 900kDa[178].

Even though much progress has been made regarding the structure determination of membrane proteins by NMR, still only quite a few examples of complete NMR structures are available, for example for the subunit c of the F_1,F_0-ATPase[179], the double transmembrane domain of the bacterial mercury transport membrane protein[180] and the human glycine receptor[181], all of which comprise two transmembrane domains. No structural information on more than one transmembrane domain from a GPCR as derived from NMR measurements has been published.

1.4.3 Membrane Mimetics for NMR Studies

Membrane proteins, also termed integral membrane proteins, can only exert their functions when they are inserted in the membrane. However naturally occurring membranes are characterized by the following features: they are patchy, with segregated regions of structure and function, and lipid regions also vary in thickness and composition(Figure 16)[182].

Therefore it is the intrinsic inhomogeneity of the membranes that confers one of the largest challenges for the *in vitro* studies of membrane proteins, in that most of techniques used today for structural studies such like X-ray crystallography and NMR need an isotropic medium to increase the signal strength. To circumvent this problem, various alternative media have been applied: organic solvents, micelles, bicelles, amphipols, nanoscale bilayes and reverse micelles.

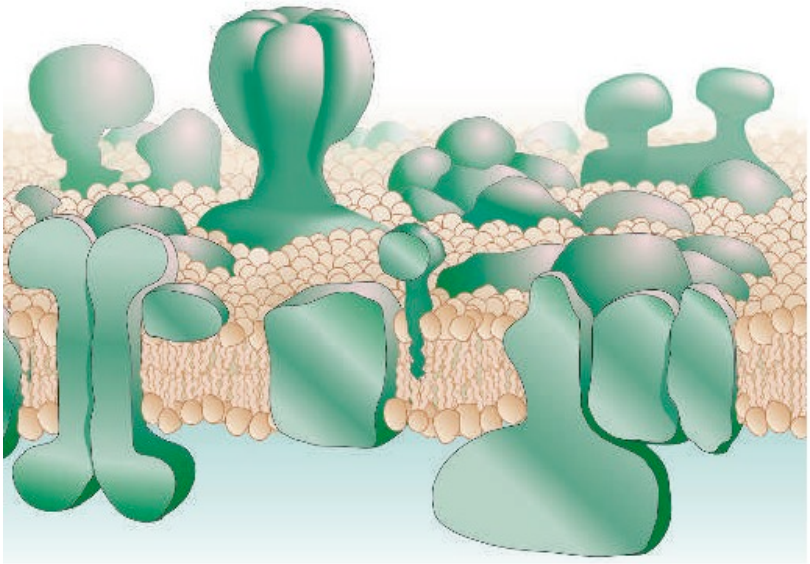


Figure 16. Mosaic membrane

Historically, organic solvents have been extensively used to handle small hydrophobic compounds, therefore they are naturally options for hydrophobic biomolecules such as membrane proteins (Figure 17d). The most often used organic solvents include DMSO, 4 : 4 : 1 chloroform : methanol : water and 80 : 20 hexafluoroisopropanol : water. Katragadda and coworkers have solved all the transmembrane domains from bacteriorhodopsin in DMSO and the resolved structures can be reasonalbly overlapped with the known crystal structure, which confirms the 2-step membrane protein folding model and indicates the applicability of organic

solvents in NMR studies[176]. In another example, the subunit c of F_1,F_0-ATPase, has also been resolved to atomic level in a 4 : 4 : 1 chloroform : methanol : water mixture[179], a solvent system in which activity is retained completely. However organic solvents are not always a good choice for membrane proteins, since many other membrane proteins have shown to be denatured in these solvents or loss of functions occurred. Looking at the two successful examples it seems that both comprise almost exclusively hydrophobic residues, on the contrary in most membrane proteins the domains out of membrane are additionally composed of hydrophilic residues, and their architecture is adjusted to the gradient in hydrophobicity encountered in biological membranes. Thus when placed in the isotropic organic environment they become leading to further conformational changes or loss of function. In summary, organic solvents are rarely a good choice for membrane proteins, and if used careful validation of the structure and function is necessary.

Micelles are formed by detergents. When detergents are dispersed in water, they are mainly monomeric. However when the concentration increases above the critical micelle concentration (CMC), monomeric detergent molecules start associating and form a spherical aggregate termed a micelle (shown in Figure 17b). Hydrophilic head groups point outward interacting with water while the hydrophobic tails point inward forming an apolar environment. Detergents can be classified into the following categories according to charge: non-ionic (sugar-derived detergents) and ionic detergents. Among ionic detergents there are cationic, anionic (SDS) and zwitterionic (CHAPS, DPC) detergents. Detergent micelles are by far the most widely used membrane mimetic for solution NMR. The transmembrane domain of glycophorin A was determined in DPC micelles in dimeric form, and the atomic structure indicates that van der Waals interaction is responsible for dimerization[169]. Howell, et al have determined the structure of the double transmembrane domain of bacterial mercury transporter in SDS[180]. DHPC was applied for the structural determination of outer membrane protein X (OmpX) and it renders the first 3-D structure of beta barrel membrane protein determined by NMR[183]. Roosild and coworkers have determined the 3-D structure of a membrane-integrating sequence for translation of integral membrane protein constructs (MISTIC) in LDAO. The structure demonstrated a helical bundle with a lipid-facing surface that was surprisingly polar. Additional experiments suggest that MISTIC can be used for high-level production of other membrane proteins in their native conformations, including many eukaryotic proteins

that have previously been intractable to bacterial expression[184]. In spite of the wide application of micelles, extra care must be paid: i) there is no "one-for-all" rule in detergent selection, therefore for each membrane protein an extensive detergent screening is needed in order to end up with one in which the native structure and hopefully the native function can be retained; ii) detergents such as SDS or others with charges right connected to the alkyl chains have stronger denaturing activity, in this concern the lysolipids are preferable; iii) the detergent concentration should be much higher than CMC and ensure that each micelle hold only one protein molecule[185].

Bicelles (shown in Figure 17a) consist of one or more long chain phospholipids which self assemble into a bilayer, in combination with an amphiphile such as short chain lipid DHPC[186]. The amphiphile serves to form a rim around the hydrophobic interface of the bilayer, creating a disk-shaped domain. The appropriate size of bicelles can be manipulated by changing the ratio between long chain and short chain lipids, therefore the small bicelles can be applied in solution NMR. Bicelles have proven to be equally useful in solid-state NMR applications, and high-quality spectra are often observed due in part to the highly liquid crystalline environment of the bicelle lipids. Bicelles have been regarded as a better membrane model than any other mimetics such like micelles: i) bicelles display a disk-like surface, thus providing a large planar area without curvature; ii) bicelles are composed of phospholipids mixtures which have proved to be critical for the function of many membrane proteins. The best example of bicelles comes from the *Staphylococcal* multidrug resistance transporter (Smr)[187]. The Girvin group has identified LPPG as the best micellar system for structural studies of IMPs according to signal dispersion in [^{15}N, ^{1}H]-HSQC spectra. In addition, equilibrium sedimentation confirmed the presence of the assembly of a dimeric state, and carbon chemical shift based secondary structure prediction indicates the presence of a 4-helix bundle. However the functional test with ligands demonstrated unspecific binding, indicating a non-native structure. Based on the functional analysis bicelles stood out as the best membrane mimetic. It is therefore believed that bicelles might be a generally useful membrane model for membrane protein studies.

Amphipols (shown in Figure 17c) are short amphipathic polymers that can substitute detergents to keep membrane proteins water-soluble. Since amphipols are not detergents, they provide an appealing advantage to avoid the denaturing or

dissociating effect of detergents. However amphipols as a non-detergent polymers can not solubilize membrane proteins directly, which means they can only be applied after the membrane proteins have been extracted by detergents. Amphipols have been for the first time used as the medium in a study leading to a high-quality NMR spectra of OmpA[188; 189], and it is interesting to see if this medium can be employed to other membrane proteins, especially to those that do not result in good spectra in conventional membrane mimetics.

Reverse micelles (shown in Figure 17e) as indicated by the name have hydrophilic head groups pointing inward and the hydrophobic tail outward. Reverse micelles are formed when the bulky solution is organic solvent and a small quantity of aqueous solution is wrapped inside. The largest advantage of reverse micelles is that the tumbling of the whole aggregate is very fast if a low viscosity organic solvent is chosen, thus leading to extraordinarily sharp lines. Reverse micelles have been successfully applied to soluble proteins and the Wand lab has been working to extend this approach to membrane proteins[190].

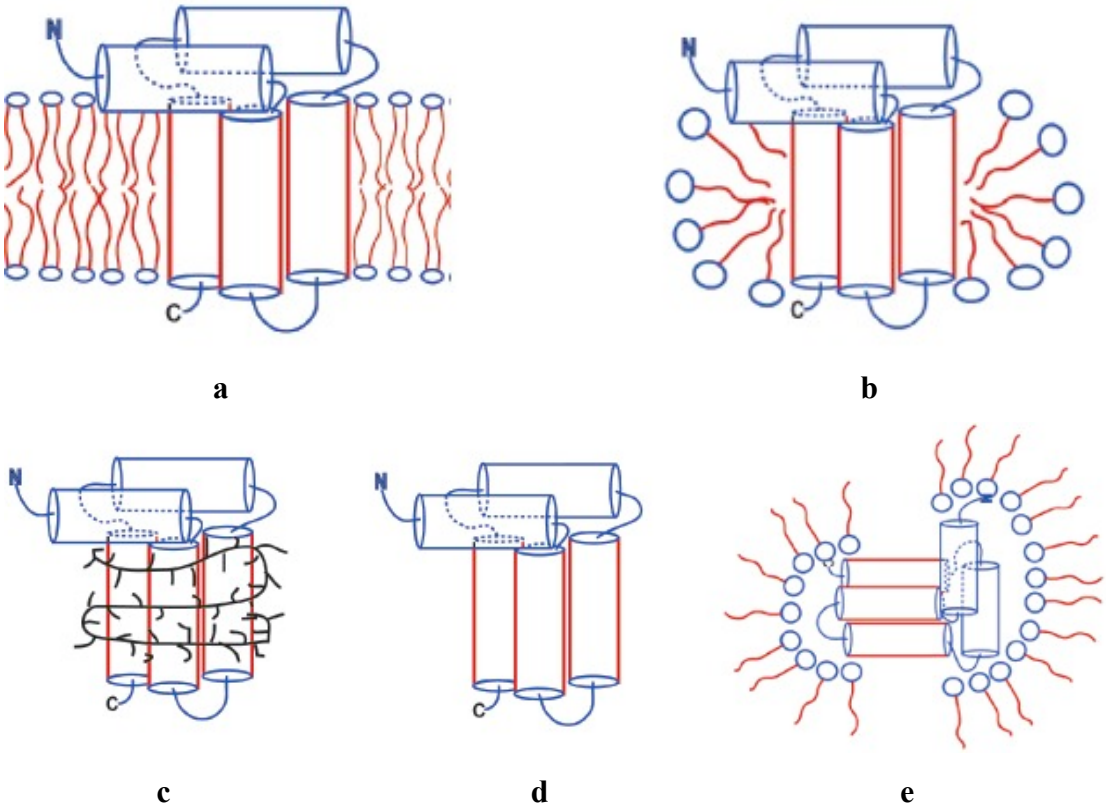


Figure 17. Membrane mimetics, red color indicates hydrophobic parts and blue color indicates hydrophilic parts. a) bicelles; b) micelles; c) amphipoles; d) organic solvent; d) reverse micelles

Nano-scale bilayers also called nano-discs are composed of a patch of planar lipid bilayers (~160 lipid molecules) surrounded by a dimmer of apolipoprotein A-I. The nano-disc has a disc-like shape with diameter of around 10-12 nm and a thickness of ~4nm. While other model membranes (bicelles, micelles) display a polydisperse particle size distribution and can suffer from problems of inconsistency and instability, nano-discs demonstrate an excellent monodispersity because the particle size is constrained by the coat of the scaffold protein(s). This system has been recently adapted by Sligar group to incorporate various membrane poteins[191], and the Arseniev group has for the first time applied this model membrane to study the structure of a peripheral membrane peptide by solution NMR[192]. Unfortunately, no results from integral membrane proteins have been reported yet.

1.5 Production of Membrane Protein Fragments

The structural determination of membrane proteins is considerably lagging behind in comparison to the soluble counterparts, and one major obstacle is the recombinant production of target membrane protein in sufficient quantity. Various expression systems have been employed for the entire membrane proteins, on the contrary for membrane fragment production *E.coli* is still the dominant choice. Another advantage of *E.coli* over other organisms with regard to NMR is easy labeling method. For other organisms the isotope labeling is always problematic due to either applicability or high cost.

1.5.1 The Fusion Method

The first method is the one, which is used most widely for 2-TM constructs. In brief, the target protein is expressed as a fusion protein with a partner, which usually can be expressed to a high level thus improving the yield of the target protein. The fusion partner can also provide other features to the fusion protein, on basis of which it can be divided into two categories: soluble fusions and insoluble fusions. Soluble fusions, as indicated by the name, can increase the solubility of the whole fusion protein and thus simplify the subsequent purification. In case of membrane proteins highly hydrophobic trasmembrane helices are present, the relatively large soluble fusion proteins, such like maltose binding protein (MBP) and glutathione S transferase (GST), are preferred provided that the fusion protein can be expressed in soluble form. After purification of the soluble fusion protein, removal of the soluble fusion is essential, which is in most cases accomplished by enzymatic cleavage. The problem with the soluble fusion approach is that it is difficult to predict whether the fusion protein will be soluble or not. If the fusion protein is insoluble, the enzymatic cleavage will be problematic requiring detergent screening in order to solubilize the fusion protein, and low cleavage efficiency is often observed. Insoluble fusion partners have been widely used to isolate double transmembrane domains. In such a system usually a very hydrophobic protein is used as the fusion partner such like the ketosteroidisomerase (KSI). Since the final fusion protein is more hydrophobic, it will most likely accumulate as inclusion bodies. Inclusion bodies can alleviate the

normally encountered problems like toxicity and increase the final yield. Furthermore the formation of inclusion bodies can also prevent unwanted degradation during expression. Even though the insoluble fusion is a robust method, a magnificent problem is related to the cleavage of the fusion protein. Since the fusion protein is not soluble in most of the enzyme-compatible solutions, cyanogens bromide in guanidine chloride/acetate acid in most cases is employed[193], which then requires that no methionine should be present in the target sequence. Hydroxylamine is another option, but its application has been severely hampered due to the poor unspecificity[67]. Another obstacle with the insoluble fusion method is that the overexpression levels are lower or even approaching undetectable levels when the target protein is getting large. In our experience in the case of the KSI system the yield of the fusion protein will decrease drastically once the target protein exceeds 100 residues.

1.5.2 Direct Expression

Direct expression is often used for the expression of large proteins, however seldom used for small ones, especially in case of double transmembrane segments, which approximately account for 10kDa. Direct expression frequently is confronted with problems such like cell toxicity, degradation and so on. However, in case of membrane protein fragments, the significant hydrophobicity will result in formation of inclusion bodies, therefore avoiding the potential cell toxicity and degradation. The non-fusion approach is seemingly complementary to the fusion method, which can not cope with large proteins. The advantage of this method is apparent: The chemical/enzymatic cleavage step is skipped, thus yields will increase and time will be saved. In addition, in case of one of our double transmembrane protein, this method has led to higher expression levels in comparison to the fusion method. The success of this method is difficult to predict.

1.5.3 Strain Selection

Although it is better to express the target protein in a closely related organism, *E.coli* is still the favorable expression host because of its rapid growth, the availability of cheap growth media, clearly defined genetics and the availability of various plasmids and mutant strains. Working with mammalian cells and yeast is more time

consuming and expensive, most importantly in case of isotope labeling essential for modern NMR studies, which requires media of defined composition often not applicable for higher organisms. The most commonly used strain is BL21 (DE3) for homologous expression as it is lon and OmpT protease deficient thus helping to stabilize the plasmid. Many derivatives on basis of BL21 (DE3) have been introduced that present additional advantages: BL21 (pLys) Rosetta contains a lysozyme gene which prevents leakage expression and contains extra copies of *E.coli* rare codons which corrects the codon usage bias; C41 (DE3) and C43 (DE3) are known to be suitable for expression of proteins which are toxic to the host; BL21-AI takes use of arabinose as the inducer, realizing extremely low leakage expression and increased resistance to toxicity.

1.5.4 Conditions for Culturing

As expression of proteins in *E.coli* is rarely trivial, modulating temperature, concentration of inducer or additives to the medium can significantly alter protein expression levels.

It is well known that lower temperatures will result in more soluble protein expressed by cells, while higher temperature is detrimental since the additional stress induced at higher temperature will lead to formation of inclusion bodies. It is also noted that the proteolysis increases at higher temperature.

Constructs controlled by inducible promoters are also subject to varying protein expression levels by altering concentrations of the inducing agent. As in the commonly used pET expression vector system, lowering IPTG concentrations may facilitate protein expression as low IPTG concentrations may cause incomplete or partial induction of protein synthesis. This reduction in protein synthesis has been associated with more reliable protein folding, subsequently improving the solubility of the product.

It has been mentioned that altering the expression medium may also aid in membrane protein expression. Supplementing media with glucose (0.2–1%) can increase protein expression, since increasing glucose concentrations is believed to decrease promoter repression resulting in improved protein synthesis. In order to obtain functional membrane proteins, the ligands for the receptors can be added.

In summary there is no generally valid approach for the preparation of membrane proteins, therefore it is always important to conduct a thorough screening through expression methods, strain selection and culture conditions in order to reach optimal yields.

1.5.5 Potential Problems

Aggregation, polymerization and conformational exchange are detrimental for NMR studies, because they will result in line-broadening, duplicated or missing signals, variability in signal strength etc. Aggregation can arise from the high hydrophobicity, which can be dealt with by varyious different detergents. The presence of cysteines is also a common cause for aggregation, especially when bivalent cations are used in the purification, such like Ni+ in the affinity purification. Of course, addition of a reducing agent is the easiest way to eliminate such effect[194]. It has been observed that the aggregation is mainly mediated by special motifs present in the transmembrane domains, such like GXXXG[195] and QXXS[196]. In the study of the rhodopsin receptor by solution NMR, it was observed that the backbone is in conformational exchange on the mili-to-micro second time scale [171; 172], however in some cases the conformational exchange will fall into a regime where NMR is not applicable any longer[197; 198]. In such a case, different detergents should be tried in order to manipulate the exchange rate.

1.6 Summary of the Work Described in this Thesis

The following chapters describe my Ph.D. work in logic order: chapter 2: “Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family”; chapter 3: “Properties of N-terminal domains from Y receptors probed by NMR spectroscopy”; chapter 4: “Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a human GPCR”, respectively.

In chapter 2 I have tried studied the structure and function of the N terminus from the Y4 (N-Y4) receptor in order to test the hypothesis that peptide hormones of the NPY family bind to their G-protein coupled receptors via forming contacts to the N-terminal domain in addition to established contacts to the extracellular loops. Therefore, I have produced the 41-mer peptide corresponding to the N-terminal domain via a recombinant method. Since initial trials for producing the target peptide with a soluble fusion partner led to unspecific degradation during expression, a novel method was employed. In brief, an insoluble fusion, ketosteroidisomerase (KSI), was used as the fusion partner which then results in accumulation in inclusion bodies and consequently avoids undesired degradation. Thereafter a detergent screening was conducted and sarcosyl, an anionic detergent, was finally chosen because it can solubilize the inclusion bodies and at the same time retain activity of the TEV protease that is used for the subsequent cleavage of the fusion protein. Following liberation from the fusion partner by TEV cleavage and conventional HPLC purification, sufficient quantities of the target peptide with good purity can be obtained. This recombinant method was for the first time used in our lab and we believe it can serve as a generic approach for the production of small peptides, in particular those that are soluble and susceptible to the unspecific cleavage during expression. The structural studies were carried out by solution NMR both in absence and in presence of membrane mimicking micelles using the ^{15}N-labelled peptide. The peptide is unstructured in solution whereas a micelle-associated helical segment is formed in the presence of dodecylphosphocholine (DPC) or sodiumdodecylsulfate (SDS) micelles. The induced secondary structure in both DPC and SDS is identical, which suggests that induction of secondary structure is regulated mainly by hydrophobic interactions instead of the charged head groups of the detergents. The

hydrophobicity of the target sequence (poly-leucine) is consistent with this conclusion. A 12-carbon lipid chain was coupled to the C terminus of the peptide, which anchors the peptide on the membrane surface and mimics the first transmembrane domain. The high similarity of spectra recorded with N-Y4 and the lipid-modified N-Y4 demonstrates that the structure of the N-terminal domain from the Y4 receptor does not depend on the presence of the first transmembrane. As measured by surface-plasmon resonance spectroscopy (SPR) N-Y4 binds with approximately 50μM affinity to the pancreatic polypeptide (PP), a high-affinity ligand at the Y4 receptor, whereas binding to neuropeptide Y (NPY) and peptide YY (PYY) is much weaker. Using site-directed mutagenesis residues in PP and in N-Y4 critical for binding are identified. The data indicate that electrostatic interactions dominate and that acidic ligand and basic receptor residues are mediating this interaction. Residues of N-Y4 are likely to contribute to the binding of PP, and in addition may possibly also help to transfer the hormone from the membrane-bound state into the receptor binding pocket.

In chapter 3 the production of all other N termini of Y receptors, namely N-Y1, N-Y2, N-Y5 is described. Similar to N-Y4, N-Y1 was expressed as an insoluble fusion followed by TEV cleavage in presence of detergent, whereas N-Y2 and N-Y5 were expressed as soluble fusions to ubiquitin followed by cleavage with ubiquitin hydrolase in aqueous solution. All N-terminal domains including N-Y4 are fully flexible in aqueous buffer. The presence of phospholipid micelles rigidifies the the peptides to various extent, with N-Y2 remaining flexible, and N-Y1 and N-Y4 becoming partially helically folded. Using chemical shift mapping techniques interactions of NPY, PYY and PP, the three members of the neurohormone family which are the Y receptors' natural ligands, with N-Y1, N-Y2 and N-Y5 revealed chemical shift changes in all cases, with the largest values being encountered for PP interacting with N-Y1 or N-Y5. The strength of the interactions, however, is generally weak, and the data also point to non-specific electrostatic contacts.

In chapter 4 on basis of N-Y4 a fragment from the Y4 receptor, that comprises the N-terminal domain, the first two transmembrane (TM) helices and the first extracellular loop followed by a $(His)_6$ tag was recombinantly expressed and purified. The initial production trials as fusions (soluble and insoluble fusion) resulted in low overexpression levels, and therefore I optimized direct expression. To my knowledge

direct expression for the first time was applied to a membrane protein fragment. It yielded 6 mg protein after the final purification step. Two points were crucial in order to obtain the chemically and conformationally pure sample: i) a large quantity of reducing agent has to be added during purification to remove the aggregates formed by unspecific disulfide linkages; ii) RP-HPLC is indispensable to remove the SDS-PAGE-invisible non-proteineous contaminants. Extensive detergent screening was required to yield good spectra, and only a combination of detergents provided optimal sample conditions. A detergent mixture 1% dodecylphosphocholine (DPC) / 6% 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG) was selected finally. Therein, DPC provided sufficient solubility for the protein whereas LPPG helped improving spectral resolution. Under optimized conditions (detergent mixture, pH and temperature) almost complete assignment of the, including all resonances from the TM segments is possible. Data on internal backbone dynamics revealed that the fragment folded predominantly into secondary structure, which was further confirmed by CD spectroscopy. The presence of the TM helices was established based on secondary chemical shifts and sequential amide proton nuclear Overhauser effects. Interestingly, the properties of the N-terminal domain in this large fragment are highly similar to those determined by us previously on the isolated N-terminal domain in the presence of DPC micelles. This supports our notion that the structure of N-Y4 is formed independent from the presence of the first TM domain. Structural studies of GPCRs have been largely delayed by the intrinsic difficulties involved in the expression, purification and stabilization in membrane mimetics of this pharmaceutically important protein superfamily. Nonetheless, the 2-step model from Engelman has inspired structural studies using fragments of membrane proteins. In my thesis I have purified all the N-terminal domains from Y receptors. In addition to the structural studies, which suggest a significant effect of the membrane on the structure of these otherwise unstructured peptides, the binding studies of N-Y4 revealed a specific recognition towards its native ligand bPP. This observation indicates that the N-terminal domain of the Y4 receptor might play a role in ligand recognition or in the ligand transfer to the receptor binding pocket. Though 2-TM domains have attracted increasing attention amongst structural biologists, so far only a handful of NMR structures are available: F_1,F_0-ATPase[179], the bacterial mercury transport membrane protein[180] and the human glycine receptor[181]. It seems that production of 2-TM domains, though only a fragment of the entire GPCR, is difficult.

The second part of my thesis describes ways how to express and purify a 2-TM fragment from the Y4 receptor. It also develops a method to optimize the detergent, and ends with a procedure that establishes sample conditions under which almost complete backbone assignment can be achieved. First of all, direct expression, which has been widely used to express the entire protein, is shown to be applicable to GPCR fragments. Secondly, the usage of a detergent mixture leads to much superior spectra, increases the choices of membrane mimetics by taking advantages of the individual properties of various detergents or lipids. The 2-TM fragment to our knowledge is the only fragment from a human GPCR described, and it is therefore of much interest to solve the structure of such a construct. I believe we have made substantial progress to achieve such an ambitious goal.

1.7 References

1. Stevens, T. J. & Arkin, I. T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* **39**, 417-20.
2. Boyd, D., Schierle, C. & Beckwith, J. (1998). How many membrane proteins are there? *Protein Sci* **7**, 201-5.
3. von Heijne, G. (1999). Recent advances in the understanding of membrane protein assembly and structure. *Q Rev Biophys* **32**, 285-307.
4. White, S. H. & Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319-65.
5. Schleiff, E. & Soll, J. (2005). Membrane protein insertion: mixing eukaryotic and prokaryotic concepts. *EMBO Rep* **6**, 1023-7.
6. White, S. H. (2004). The progress of membrane protein structure determination. *Protein Sci* **13**, 1948-9.
7. Foord, S. M. (2002). Receptor classification: post genome. *Curr Opin Pharmacol* **2**, 561-6.
8. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science* **291**, 1304-51.
9. Jacoby, E., Bouhelal, R., Gerspacher, M. & Seuwen, K. (2006). The 7 TM G-protein-coupled receptor target family. *ChemMedChem* **1**, 761-82.
10. Hopkins, A. L. & Groom, C. R. (2002). The druggable genome. *Nat Rev Drug Discov* **1**, 727-30.
11. Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F. E. & Vriend, G. (2003). GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res* **31**, 294-7.
12. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **63**, 1256-72.
13. Lagerstrom, M. C. & Schioth, H. B. (2008). Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* **7**, 339-57.

14. Harmar, A. J. (2001). Family-B G-protein-coupled receptors. *Genome Biol* **2**, REVIEWS3013.
15. Baneres, J. L., Martin, A., Hullot, P., Girard, J. P., Rossi, J. C. & Parello, J. (2003). Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1. *J Mol Biol* **329**, 801-14.
16. Kiefer, H., Vogel, R. & Maier, K. (2000). Bacterial expression of G-protein-coupled receptors: prediction of expression levels from sequence. *Receptors Channels* **7**, 109-19.
17. Weiss, H. M. & Grisshammer, R. (2002). Purification and characterization of the human adenosine A(2a) receptor functionally expressed in Escherichia coli. *Eur J Biochem* **269**, 82-92.
18. Grisshammer, R., White, J. F., Trinh, L. B. & Shiloach, J. (2005). Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview. *J Struct Funct Genomics* **6**, 159-63.
19. Sarramegna, V., Talmont, F., Demange, P. & Milon, A. (2003). Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification. *Cell Mol Life Sci* **60**, 1529-46.
20. Klammt, C., Schwarz, D., Eifler, N., Engel, A., Piehler, J., Haase, W., Hahn, S., Dotsch, V. & Bernhard, F. (2007). Cell-free production of G protein-coupled receptors for functional and structural studies. *J Struct Biol* **158**, 482-93.
21. Eroglu, C., Cronet, P., Panneels, V., Beaufils, P. & Sinning, I. (2002). Functional reconstitution of purified metabotropic glutamate receptor expressed in the fly eye. *EMBO Rep* **3**, 491-6.
22. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-45.
23. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. (2008). Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*.
24. Murakami, M. & Kouyama, T. (2008). Crystal structure of squid rhodopsin. *Nature* **453**, 363-7.
25. Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Yao, X. J., Weis, W. I., Stevens, R. C. & Kobilka, B. K. (2007). GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**, 1266-73.
26. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-65.
27. Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C., Henderson, R., Leslie, A. G., Tate, C. G. & Schertler, G. F. (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-91.
28. Tian, C., Breyer, R. M., Kim, H. J., Karra, M. D., Friedman, D. B., Karpay, A. & Sanders, C. R. (2005). Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor. *J Am Chem Soc* **127**, 8010-1.

29. Gautier, A., Kirkpatrick, J. P. & Nietlispach, D. (2008). Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary structure, and dynamics. *Angew Chem Int Ed Engl* **47**, 7297-300.
30. Werner, K., Richter, C., Klein-Seetharaman, J. & Schwalbe, H. (2008). Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* **40**, 49-53.
31. Maggio, R., Vogel, Z. & Wess, J. (1993). Reconstitution of functional muscarinic receptors by co-expression of amino- and carboxyl-terminal receptor fragments. *FEBS Lett* **319**, 195-200.
32. Hebert, T. E., Moffett, S., Morello, J. P., Loisel, T. P., Bichet, D. G., Barret, C. & Bouvier, M. (1996). A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J Biol Chem* **271**, 16384-92.
33. Marshall, F. H., Jones, K. A., Kaupmann, K. & Bettler, B. (1999). GABAB receptors - the first 7TM heterodimers. *Trends Pharmacol Sci* **20**, 396-9.
34. Bulenger, S., Marullo, S. & Bouvier, M. (2005). Emerging role of homo- and heterodimerization in G-protein-coupled receptor biosynthesis and maturation. *Trends Pharmacol Sci* **26**, 131-7.
35. Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A. & Palczewski, K. (2003). Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* **421**, 127-8.
36. George, S. R., Fan, T., Xie, Z., Tse, R., Tam, V., Varghese, G. & O'Dowd, B. F. (2000). Oligomerization of mu- and delta-opioid receptors. Generation of novel functional properties. *J Biol Chem* **275**, 26128-35.
37. Gomes, I., Gupta, A., Filipovska, J., Szeto, H. H., Pintar, J. E. & Devi, L. A. (2004). A role for heterodimerization of mu and delta opiate receptors in enhancing morphine analgesia. *Proc Natl Acad Sci U S A* **101**, 5135-9.
38. Ribas, C., Penela, P., Murga, C., Salcedo, A., Garcia-Hoz, C., Jurado-Pueyo, M., Aymerich, I. & Mayor, F., Jr. (2007). The G protein-coupled receptor kinase (GRK) interactome: role of GRKs in GPCR regulation and signaling. *Biochim Biophys Acta* **1768**, 913-22.
39. Shenoy, S. K. & Lefkowitz, R. J. (2005). Seven-transmembrane receptor signaling through beta-arrestin. *Sci STKE* **2005**, cm10.
40. Morii, H. & Watanabe, Y. (1992). A possible role of carbohydrate moieties in prostaglandin D2 and prostaglandin E2 receptor proteins from the porcine temporal cortex. *Arch Biochem Biophys* **292**, 121-7.
41. Nakagawa, M., Miyamoto, T., Kusakabe, R., Takasaki, S., Takao, T., Shichida, Y. & Tsuda, M. (2001). O-Glycosylation of G-protein-coupled receptor, octopus rhodopsin. Direct analysis by FAB mass spectrometry. *FEBS Lett* **496**, 19-24.
42. Innamorati, G., Sadeghi, H. & Birnbaumer, M. (1996). A fully active nonglycosylated V2 vasopressin receptor. *Mol Pharmacol* **50**, 467-73.
43. Russo, D., Chazenbalk, G. D., Nagayama, Y., Wadsworth, H. L. & Rapoport, B. (1991). Site-directed mutagenesis of the human thyrotropin receptor: role of asparagine-linked oligosaccharides in the expression of a functional receptor. *Mol Endocrinol* **5**, 29-33.
44. Dattatreyamurty, B. & Reichert, L. E., Jr. (1992). Carbohydrate moiety of follitropin receptor is not required for high affinity hormone-binding or for

functional coupling between receptor and guanine nucleotide-binding protein in bovine calf testis membranes. *Endocrinology* **131**, 2437-45.
45. Rens-Domiano, S. & Reisine, T. (1991). Structural analysis and functional role of the carbohydrate component of somatostatin receptors. *J Biol Chem* **266**, 20094-102.
46. Kaushal, S., Ridge, K. D. & Khorana, H. G. (1994). Structure and function in rhodopsin: the role of asparagine-linked glycosylation. *Proc Natl Acad Sci U S A* **91**, 4024-8.
47. Rands, E., Candelore, M. R., Cheung, A. H., Hill, W. S., Strader, C. D. & Dixon, R. A. (1990). Mutational analysis of beta-adrenergic receptor glycosylation. *J Biol Chem* **265**, 10759-64.
48. Benya, R. V., Fathi, Z., Kusui, T., Pradhan, T., Battey, J. F. & Jensen, R. T. (1994). Gastrin-releasing peptide receptor-induced internalization, down-regulation, desensitization, and growth: possible role for cyclic AMP. *Mol Pharmacol* **46**, 235-45.
49. Fukushima, Y., Oka, Y., Saitoh, T., Katagiri, H., Asano, T., Matsuhashi, N., Takata, K., van Breda, E., Yazaki, Y. & Sugano, K. (1995). Structural and functional analysis of the canine histamine H2 receptor by site-directed mutagenesis: N-glycosylation is not vital for its action. *Biochem J* **310 (Pt 2)**, 553-8.
50. Piersen, C. E., True, C. D. & Wells, J. N. (1994). A carboxyl-terminally truncated mutant and nonglycosylated A2a adenosine receptors retain ligand binding. *Mol Pharmacol* **45**, 861-70.
51. Servant, G., Dudley, D. T., Escher, E. & Guillemette, G. (1996). Analysis of the role of N-glycosylation in cell-surface expression and binding properties of angiotensin II type-2 receptor of rat pheochromocytoma cells. *Biochem J* **313 (Pt 1)**, 297-304.
52. Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M. & Aebi, M. (2006). Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* **25**, 1957-66.
53. Karnik, S. S., Ridge, K. D., Bhattacharya, S. & Khorana, H. G. (1993). Palmitoylation of bovine opsin and its cysteine mutants in COS cells. *Proc Natl Acad Sci U S A* **90**, 40-4.
54. Fukushima, Y., Saitoh, T., Anai, M., Ogihara, T., Inukai, K., Funaki, M., Sakoda, H., Onishi, Y., Ono, H., Fujishiro, M., Ishikawa, T., Takata, K., Nagai, R., Omata, M. & Asano, T. (2001). Palmitoylation of the canine histamine H2 receptor occurs at Cys(305) and is important for cell surface targeting. *Biochim Biophys Acta* **1539**, 181-91.
55. Percherancier, Y., Planchenault, T., Valenzuela-Fernandez, A., Virelizier, J. L., Arenzana-Seisdedos, F. & Bachelerie, F. (2001). Palmitoylation-dependent control of degradation, life span, and membrane expression of the CCR5 receptor. *J Biol Chem* **276**, 31936-44.
56. Gao, Z., Ni, Y., Szabo, G. & Linden, J. (1999). Palmitoylation of the recombinant human A1 adenosine receptor: enhanced proteolysis of palmitoylation-deficient mutant receptors. *Biochem J* **342 (Pt 2)**, 387-95.
57. O'Brien, P. J. & Zatz, M. (1984). Acylation of bovine rhodopsin by [3H]palmitic acid. *J Biol Chem* **259**, 5054-7.
58. O'Dowd, B. F., Hnatowich, M., Caron, M. G., Lefkowitz, R. J. & Bouvier, M. (1989). Palmitoylation of the human beta 2-adrenergic receptor. Mutation of

Cys341 in the carboxyl tail leads to an uncoupled nonpalmitoylated form of the receptor. *J Biol Chem* **264**, 7564-9.
59. Chen, C., Shahabi, V., Xu, W. & Liu-Chen, L. Y. (1998). Palmitoylation of the rat mu opioid receptor. *FEBS Lett* **441**, 148-52.
60. Monteclaro, F. S. & Charo, I. F. (1997). The amino-terminal domain of CCR2 is both necessary and sufficient for high affinity binding of monocyte chemoattractant protein 1. Receptor activation by a pseudo-tethered ligand. *J Biol Chem* **272**, 23186-90.
61. Michel, M. C., Beck-Sickinger, A., Cox, H., Doods, H. N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T. & Westfall, T. (1998). XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors. *Pharmacol Rev* **50**, 143-50.
62. Gregor, P., Feng, Y., DeCarr, L. B., Cornfield, L. J. & McCaleb, M. L. (1996). Molecular characterization of a second mouse pancreatic polypeptide receptor and its inactivated human homologue. *J Biol Chem* **271**, 27776-81.
63. Larhammar, D. & Salaneck, E. (2004). Molecular evolution of NPY receptor subtypes. *Neuropeptides* **38**, 141-51.
64. Bromee, T., Sjodin, P., Fredriksson, R., Boswell, T., Larsson, T. A., Salaneck, E., Zoorob, R., Mohell, N. & Larhammar, D. (2006). Neuropeptide Y-family receptors Y6 and Y7 in chicken. Cloning, pharmacological characterization, tissue distribution and conserved synteny with human chromosome region. *FEBS J* **273**, 2048-63.
65. Larhammar, D., Wraith, A., Berglund, M. M., Holmberg, S. K. & Lundell, I. (2001). Origins of the many NPY-family receptors in mammals. *Peptides* **22**, 295-307.
66. Berglund, M. M., Lundell, I., Eriksson, H., Soll, R., Beck-Sickinger, A. G. & Larhammar, D. (2001). Studies of the human, rat, and guinea pig Y4 receptors using neuropeptide Y analogues and two distinct radioligands. *Peptides* **22**, 351-6.
67. Cabrele, C., Wieland, H. A., Langer, M., Stidsen, C. E. & Beck-Sickinger, A. G. (2001). Y-receptor affinity modulation by the design of pancreatic polypeptide/neuropeptide Y chimera led to Y(5)-receptor ligands with picomolar affinity. *Peptides* **22**, 365-78.
68. McCrea, K., Wisialowski, T., Cabrele, C., Church, B., Beck-Sickinger, A., Kraegen, E. & Herzog, H. (2000). 2-36[K4,RYYSA(19-23)]PP a novel Y5-receptor preferring ligand with strong stimulatory effect on food intake. *Regul Pept* **87**, 47-58.
69. Pedrazzini, T., Seydoux, J., Kunstner, P., Aubert, J. F., Grouzmann, E., Beermann, F. & Brunner, H. R. (1998). Cardiovascular response, feeding behavior and locomotor activity in mice lacking the NPY Y1 receptor. *Nat Med* **4**, 722-6.
70. Thiele, T. E., Koh, M. T. & Pedrazzini, T. (2002). Voluntary alcohol consumption is controlled via the neuropeptide Y Y1 receptor. *J Neurosci* **22**, RC208.
71. Wheway, J., Herzog, H. & Mackay, F. (2007). NPY and receptors in immune and inflammatory diseases. *Curr Top Med Chem* **7**, 1743-52.
72. Batterham, R. L., Cowley, M. A., Small, C. J., Herzog, H., Cohen, M. A., Dakin, C. L., Wren, A. M., Brynes, A. E., Low, M. J., Ghatei, M. A., Cone, R.

D. & Bloom, S. R. (2002). Gut hormone PYY(3-36) physiologically inhibits food intake. *Nature* **418**, 650-4.

73. Malmstrom, R. E., Hokfelt, T., Bjorkman, J. A., Nihlen, C., Bystrom, M., Ekstrand, A. J. & Lundberg, J. M. (1998). Characterization and molecular cloning of vascular neuropeptide Y receptor subtypes in pig and dog. *Regul Pept* **75-76**, 55-70.
74. Zukowska-Grojec, Z., Karwatowska-Prokopczuk, E., Rose, W., Rone, J., Movafagh, S., Ji, H., Yeh, Y., Chen, W. T., Kleinman, H. K., Grouzmann, E. & Grant, D. S. (1998). Neuropeptide Y: a novel angiogenic factor from the sympathetic nerves and endothelium. *Circ Res* **83**, 187-95.
75. Naveilhan, P., Hassani, H., Canals, J. M., Ekstrand, A. J., Larefalk, A., Chhajlani, V., Arenas, E., Gedda, K., Svensson, L., Thoren, P. & Ernfors, P. (1999). Normal feeding behavior, body weight and leptin response require the neuropeptide Y Y2 receptor. *Nat Med* **5**, 1188-93.
76. Sainsbury, A., Schwarzer, C., Couzens, M., Fetissov, S., Furtinger, S., Jenkins, A., Cox, H. M., Sperk, G., Hokfelt, T. & Herzog, H. (2002). Important role of hypothalamic Y2 receptors in body weight regulation revealed in conditional knockout mice. *Proc Natl Acad Sci U S A* **99**, 8938-43.
77. Schober, D. A., Van Abbema, A. M., Smiley, D. L., Bruns, R. F. & Gehlert, D. R. (1998). The neuropeptide Y Y1 antagonist, 1229U91, a potent agonist for the human pancreatic polypeptide-preferring (NPY Y4) receptor. *Peptides* **19**, 537-42.
78. Ishihara, A., Kanatani, A., Mashiko, S., Tanaka, T., Hidaka, M., Gomori, A., Iwaasa, H., Murai, N., Egashira, S., Murai, T., Mitobe, Y., Matsushita, H., Okamoto, O., Sato, N., Jitsuoka, M., Fukuroda, T., Ohe, T., Guan, X., MacNeil, D. J., Van der Ploeg, L. H., Nishikibe, M., Ishii, Y., Ihara, M. & Fukami, T. (2006). A neuropeptide Y Y5 antagonist selectively ameliorates body weight gain and associated parameters in diet-induced obese mice. *Proc Natl Acad Sci U S A* **103**, 7154-8.
79. Marsh, D. J., Hollopeter, G., Kafer, K. E. & Palmiter, R. D. (1998). Role of the Y5 neuropeptide Y receptor in feeding and obesity. *Nat Med* **4**, 718-21.
80. Kitlinska, J., Abe, K., Kuo, L., Pons, J., Yu, M., Li, L., Tilan, J., Everhart, L., Lee, E. W., Zukowska, Z. & Toretsky, J. A. (2005). Differential effects of neuropeptide Y on the growth and vascularization of neural crest-derived tumors. *Cancer Res* **65**, 1719-28.
81. Korner, M., Waser, B. & Reubi, J. C. (2004). High expression of neuropeptide y receptors in tumors of the human adrenal gland and extra-adrenal paraganglia. *Clin Cancer Res* **10**, 8426-33.
82. Korner, M., Waser, B. & Reubi, J. C. (2004). Neuropeptide Y receptor expression in human primary ovarian neoplasms. *Lab Invest* **84**, 71-80.
83. Reubi, J. C., Korner, M., Waser, B., Mazzucchelli, L. & Guillou, L. (2004). High expression of peptide receptors as a novel target in gastrointestinal stromal tumours. *Eur J Nucl Med Mol Imaging* **31**, 803-10.
84. Korner, M. & Reubi, J. C. (2007). NPY receptors in human cancer: A review of current knowledge. *Peptides* **28**, 419-25.
85. Kamiji, M. M. & Inui, A. (2007). Neuropeptide y receptor selective ligands in the treatment of obesity. *Endocr Rev* **28**, 664-84.
86. Cabrele, C., Langer, M., Bader, R., Wieland, H. A., Doods, H. N., Zerbe, O. & Beck-Sickinger, A. G. (2000). The first selective agonist for the neuropeptide Y Y5 receptor increases food intake in rats. *J Biol Chem* **275**, 36043-8.

87. Balasubramaniam, A., Mullins, D. E., Lin, S., Zhai, W., Tao, Z., Dhawan, V. C., Guzzi, M., Knittel, J. J., Slack, K., Herzog, H. & Parker, E. M. (2006). Neuropeptide Y (NPY) Y4 receptor selective agonists based on NPY(32-36): development of an anorectic Y4 receptor selective agonist with picomolar affinity. *J Med Chem* **49**, 2661-5.
88. Sautel, M., Martinez, R., Munoz, M., Peitsch, M. C., Beck-Sickinger, A. G. & Walker, P. (1995). Role of a hydrophobic pocket of the human Y1 neuropeptide Y receptor in ligand binding. *Mol Cell Endocrinol* **112**, 215-22.
89. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. (1994). Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding. *J Biol Chem* **269**, 2863-9.
90. Nakamura, M., Sakanaka, C., Aoki, Y., Ogasawara, H., Tsuji, T., Kodama, H., Matsumoto, T., Shimizu, T. & Noma, M. (1995). Identification of two isoforms of mouse neuropeptide Y-Y1 receptor generated by alternative splicing. Isolation, genomic structure, and functional expression of the receptors. *J Biol Chem* **270**, 30102-10.
91. Fuhlendorff, J., Gether, U., Aakerlund, L., Langeland-Johansen, N., Thogersen, H., Melberg, S. G., Olsen, U. B., Thastrup, O. & Schwartz, T. W. (1990). [Leu31, Pro34]neuropeptide Y: a specific Y1 receptor agonist. *Proc Natl Acad Sci U S A* **87**, 182-6.
92. Potter, E. K., Fuhlendorff, J. & Schwartz, T. W. (1991). [Pro34]neuropeptide Y selectively identifies postjunctional-mediated actions of neuropeptide Y in vivo in rats and dogs. *Eur J Pharmacol* **193**, 15-9.
93. Balasubramaniam, A., Sheriff, S., Johnson, M. E., Prabhakaran, M., Huang, Y., Fischer, J. E. & Chance, W. T. (1994). [D-TRP32]neuropeptide Y: a competitive antagonist of NPY in rat hypothalamus. *J Med Chem* **37**, 811-5.
94. Lindner, D., van Dieck, J., Merten, N., Morl, K., Gunther, R., Hofmann, H. J. & Beck-Sickinger, A. G. (2008). GPC receptors and not ligands decide the binding mode in neuropeptide Y multireceptor/multiligand system. *Biochemistry* **47**, 5905-14.
95. Munch, G., Walker, P., Shine, J. & Herzog, H. (1995). Ligand binding analysis of human neuropeptide Y1 receptor mutants expressed in *E. coli*. *Receptors Channels* **3**, 291-7.
96. Gaudin, P., Couvineau, A., Maoret, J. J., Rouyer-Fessard, C. & Laburthe, M. (1995). Mutational analysis of cysteine residues within the extracellular domains of the human vasoactive intestinal peptide (VIP) 1 receptor identifies seven mutants that are defective in VIP binding. *Biochem Biophys Res Commun* **211**, 901-8.
97. Laburthe, M., Couvineau, A. & Marie, J. C. (2002). VPAC receptors for VIP and PACAP. *Receptors Channels* **8**, 137-53.
98. Sun, C., Song, D., Davis-Taber, R. A., Barrett, L. W., Scott, V. E., Richardson, P. L., Pereda-Lopez, A., Uchic, M. E., Solomon, L. R., Lake, M. R., Walter, K. A., Hajduk, P. J. & Olejniczak, E. T. (2007). Solution structure and mutational analysis of pituitary adenylate cyclase-activating polypeptide binding to the extracellular domain of PAC1-RS. *Proc Natl Acad Sci U S A* **104**, 7875-80.
99. Cheng, Y., Legall, T., Oldfield, C. J., Mueller, J. P., Van, Y. Y., Romero, P., Cortese, M. S., Uversky, V. N. & Dunker, A. K. (2006). Rational drug design via intrinsically disordered protein. *Trends Biotechnol* **24**, 435-42.

100. Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem Sci* **27**, 527-33.
101. Sigalov, A. B., Aivazian, D. A., Uversky, V. N. & Stern, L. J. (2006). Lipid-Binding Activity of Intrinsically Unstructured Cytoplasmic Domains of Multichain Immune Recognition Receptor Signaling Subunits. *Biochemistry* **45**, 15731-15739.
102. Ho, H. H., Du, D. & Gershengorn, M. C. (1999). The N terminus of Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor is necessary for high affinity chemokine binding but not for constitutive activity. *J Biol Chem* **274**, 31327-32.
103. Prado, G. N., Suetomi, K., Shumate, D., Maxwell, C., Ravindran, A., Rajarathnam, K. & Navarro, J. (2007). Chemokine signaling specificity: essential role for the N-terminal domain of chemokine receptors. *Biochemistry* **46**, 8961-8.
104. Andersson, H., D'Antona, A. M., Kendall, D. A., Von Heijne, G. & Chin, C. N. (2003). Membrane assembly of the cannabinoid receptor 1: impact of a long N-terminal tail. *Mol Pharmacol* **64**, 570-7.
105. Dong, M., Ding, X. Q., Thomas, S. E., Gao, F., Lam, P. C., Abagyan, R. & Miller, L. J. (2007). Role of Lysine(187) within the Second Extracellular Loop of the Type A Cholecystokinin Receptor in Agonist-Induced Activation. Use of Complementary Charge-Reversal Mutagenesis To Define a Functionally Important Interdomain Interaction. *Biochemistry* **46,** 4522-31.
106. Gupta, A., Decaillot, F. M., Gomes, I., Tkalych, O., Heimann, A. S., Ferro, E. S. & Devi, L. A. (2007). Conformation state-sensitive antibodies to G-protein-coupled receptors. *J Biol Chem* **282**, 5116-24.
107. Larhammar, D. (1996). Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **62**, 1-11.
108. Gray, T. S. & Morley, J. E. (1986). Neuropeptide Y: anatomical distribution and possible function in mammalian nervous system. *Life Sci* **38**, 389-401.
109. Ruscica, M., Dozio, E., Motta, M. & Magni, P. (2007). Relevance of the neuropeptide Y system in the biology of cancer progression. *Curr Top Med Chem* **7**, 1682-91.
110. Pedrazzini, T. (2004). Importance of NPY Y1 receptor-mediated pathways: assessment using NPY Y1 receptor knockouts. *Neuropeptides* **38**, 267-75.
111. Ekblad, E. & Sundler, F. (2002). Distribution of pancreatic polypeptide and peptide YY. *Peptides* **23**, 251-61.
112. Eberlein, G. A., Eysselein, V. E., Schaeffer, M., Layer, P., Grandt, D., Goebell, H., Niebel, W., Davis, M., Lee, T. D., Shively, J. E. & et al. (1989). A new molecular form of PYY: structural characterization of human PYY(3-36) and PYY(1-36). *Peptides* **10**, 797-803.
113. Lin, H. C. & Chey, W. Y. (2003). Cholecystokinin and peptide YY are released by fat in either proximal or distal small intestine in dogs. *Regul Pept* **114**, 131-5.
114. Schwartz, T. W. & Tager, H. S. (1981). Isolation and biogenesis of a new peptide from pancreatic islets. *Nature* **294**, 589-91.
115. Schmidt, P. T., Naslund, E., Gryback, P., Jacobsson, H., Holst, J. J., Hilsted, L. & Hellstrom, P. M. (2005). A role for pancreatic polypeptide in the regulation of gastric emptying and short-term metabolic control. *J Clin Endocrinol Metab* **90**, 5241-6.

116. Li, X. A., Sutcliffe, M. J., Schwartz, T. W. & Dobson, C. M. (1992). Sequence-specific ^{1}H NMR assignments and solution structure of bovine pancreatic polypeptide. *Biochemistry* **31**, 1245-53.
117. Frerker, N., Wagner, L., Wolf, R., Heiser, U., Hoffmann, T., Rahfeld, J. U., Schade, J., Karl, T., Naim, H. Y., Alfalah, M., Demuth, H. U. & von Horsten, S. (2007). Neuropeptide Y (NPY) cleaving enzymes: Structural and functional homologues of dipeptidyl peptidase 4. *Peptides* **28**, 257-68.
118. Grandt, D., Siewert, J., Sieburg, B., al Tai, O., Schimiczek, M., Goebell, H., Layer, P., Eysselein, V. E., Reeve, J. R., Jr. & Muller, M. K. (1995). Peptide YY inhibits exocrine pancreatic secretion in isolated perfused rat pancreas by Y1 receptors. *Pancreas* **10**, 180-6.
119. Gue, M., Junien, J. L., Reeve, J. R., Jr., Rivier, J., Grandt, D. & Tache, Y. (1996). Reversal by NPY, PYY and 3-36 molecular forms of NPY and PYY of intracisternal CRF-induced inhibition of gastric acid secretion in rats. *Br J Pharmacol* **118**, 237-42.
120. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C.-W. (1981). X-ray analysis (1.4 Å resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc Natl Acad Sci U S A* **78**, 4175-79.
121. Monks, S. A., Karagianis, G., Howlett, G. J. & Norton, R. S. (1996). Solution structure of human neuropeptide Y. *J. Biomol. NMR* **8**, 379-90.
122. Bettio, A., Dinger, M. C. & Beck-Sickinger, A. G. (2002). The neuropeptide Y monomer in solution is not folded in the pancreatic- polypeptide fold. *Protein Sci* **11**, 1834-44.
123. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. (2001).Structure and dynamics of micelle-bound neuropeptide Y: comparison with unligated NPY and implications for receptor selection. *J Mol Biol* **305**, 307-392.
124. Lerch, M., Mayrhofer, M. & Zerbe, O. (2004). Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J Mol Biol* **339**, 1153-68.
125. Lerch, M., Gafner, V., Bader, R., Christen, B., Folkers, G. & Zerbe, O. (2002). Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles. *J Mol Biol* **322**, 1117-33.
126. Seddon, A. M., Curnow, P. & Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* **1666**, 105-17.
127. Bader, R., Rytz, G., Lerch, M., Beck-Sickinger, A. G. & Zerbe, O. (2002). Key Motif to Gain Selectivity at the Neuropeptide Y5-Receptor: Solution Structure and Dynamics of [Ala31,Pro32]-NPY. *Biochemistry* **41**, 8031-42.
128. Beck-Sickinger, A. G., Wieland, H. A., Wittneben, H., Willim, K. D., Rudolf, K. & Jung, G. (1994). Complete L-alanine scan of neuropeptide Y reveals ligands binding to Y1 and Y2 receptors with distinguished conformations. *Eur J Biochem* **225**, 947-58.
129. Schwyzer, R. (1991). Peptide-membrane interactions and a new principle in quantitative structure-activity relationships. *Biopolymers* **31**, 785-92.
130. Schwyzer, R. (1995). In search of the 'bio-active conformation'--is it induced by the target cell membrane? *J. Mol. Recognit.* **8**, 3-8.
131. Kaiser, E. T. & Kezdy, F. J. (1984). Amphiphilic secondary structure: design of peptide hormones. *Science* **223**, 249-55.
132. Kaiser, E. T. & Kezdy, F. J. (1983). Secondary structures of proteins and peptides in amphiphilic environments. *Proc Natl Acad Sci U S A* **80**, 1137-43.

133. Bader, R. & Zerbe, O. (2005). Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state? *ChemBioChem* **6**, 1520-34.
134. Lerch, M., Kamimori, H., Folkers, G., Aguilar, M.-I., Beck-Sickinger, A. G. & Zerbe, O. (2005). Strongly Altered Receptor Binding Properties in PP and NPY Chimera are Accompanied by Changes in Structure and Membrane Binding. *Biochemistry* **44**, 9255 - 9264.
135. Popot, J. L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**, 4031-7.
136. Popot, J. L. & Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922.
137. Huang, K. S., Bayley, H., Liao, M. J., London, E. & Khorana, H. G. (1981). Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *J Biol Chem* **256**, 3802-9.
138. Kahn, T. W. & Engelman, D. M. (1992). Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. *Biochemistry* **31**, 6144-51.
139. Ridge, K. D., Lee, S. S. & Yao, L. L. (1995). In vivo assembly of rhodopsin from expressed polypeptide fragments. *Proc Natl Acad Sci U S A* **92**, 3204-8.
140. Martin, N. P., Leavitt, L. M., Sommers, C. M. & Dumont, M. E. (1999). Assembly of G protein-coupled receptors from fragments: identification of functional receptors with discontinuities in each of the loops connecting transmembrane segments. *Biochemistry* **38**, 682-95.
141. Wrubel, W., Stochaj, U. & Ehring, R. (1994). Construction and in vivo analysis of new split lactose permeases. *FEBS Lett* **349**, 433-8.
142. Kobilka, B. K., Kobilka, T. S., Daniel, K., Regan, J. W., Caron, M. G. & Lefkowitz, R. J. (1988). Chimeric alpha 2-,beta 2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science* **240**, 1310-6.
143. Schmidt-Rose, T. & Jentsch, T. J. (1997). Reconstitution of functional voltage-gated chloride channels from complementary fragments of CLC-1. *J Biol Chem* **272**, 20515-21.
144. Deber, C. M., Wang, C., Liu, L. P., Prior, A. S., Agrawal, S., Muskat, B. L. & Cuticchia, A. J. (2001). TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci* **10**, 212-9.
145. Moller, S., Croning, M. D. & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646-53.
146. Monne, M., Nilsson, I., Elofsson, A. & von Heijne, G. (1999). Turns in transmembrane helices: determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. *J Mol Biol* **293**, 807-14.
147. Bowie, J. U. (1997). Helix packing in membrane proteins. *J Mol Biol* **272**, 780-9.
148. Hildebrand, P. W., Preissner, R. & Frommel, C. (2004). Structural features of transmembrane helices. *FEBS Lett* **559**, 145-51.
149. von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**, 487-94.

150. Whitley, P., Nilsson, L. & von Heijne, G. (1993). Three-dimensional model for the membrane domain of Escherichia coli leader peptidase based on disulfide mapping. *Biochemistry* **32**, 8534-9.
151. Landolt-Marticorena, C., Williams, K. A., Deber, C. M. & Reithmeier, R. A. (1993). Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J Mol Biol* **229**, 602-8.
152. Hong, H., Park, S., Jimenez, R. H., Rinehart, D. & Tamm, L. K. (2007). Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. *J Am Chem Soc* **129**, 8320-7.
153. Yuen, C. T., Davidson, A. R. & Deber, C. M. (2000). Role of aromatic residues at the lipid-water interface in micelle-bound bacteriophage M13 major coat protein. *Biochemistry* **39**, 16155-62.
154. Braun, P. & von Heijne, G. (1999). The aromatic residues Trp and Phe have different effects on the positioning of a transmembrane helix in the microsomal membrane. *Biochemistry* **38**, 9778-82.
155. Partridge, A. W., Therien, A. G. & Deber, C. M. (2004). Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. *Proteins* **54**, 648-56.
156. Senes, A., Engel, D. E. & DeGrado, W. F. (2004). Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr Opin Struct Biol* **14**, 465-79.
157. Cordes, F. S., Bright, J. N. & Sansom, M. S. (2002). Proline-induced distortions of transmembrane helices. *J Mol Biol* **323**, 951-60.
158. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. (2004). The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **101**, 959-63.
159. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. & von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-81.
160. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. & von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-30.
161. Johnson, R. M., Heslop, C. L. & Deber, C. M. (2004). Hydrophobic helical hairpins: design and packing interactions in membrane environments. *Biochemistry* **43**, 14361-9.
162. Johnson, R. M., Hecht, K. & Deber, C. M. (2007). Aromatic and cation-pi interactions enhance helix-helix association in a membrane environment. *Biochemistry* **46**, 9208-14.
163. Choi, M. Y., Cardarelli, L., Therien, A. G. & Deber, C. M. (2004). Non-native interhelical hydrogen bonds in the cystic fibrosis transmembrane conductance regulator domain modulated by polar mutations. *Biochemistry* **43**, 8077-83.
164. Therien, A. G., Grant, F. E. & Deber, C. M. (2001). Interhelical hydrogen bonds in the CFTR membrane domain. *Nat Struct Biol* **8**, 597-601.
165. Monne, M., Hermansson, M. & von Heijne, G. (1999). A turn propensity scale for transmembrane helices. *J Mol Biol* **288**, 141-5.
166. Nagy, A. & Turner, R. J. (2007). The membrane integration of a naturally occurring alpha-helical hairpin. *Biochem Biophys Res Commun* **356**, 392-7.

167. Wehbi, H., Rath, A., Glibowicka, M. & Deber, C. M. (2007). Role of the extracellular loop in the folding of a CFTR transmembrane helical hairpin. *Biochemistry* **46**, 7099-106.
168. Hermansson, M., Monne, M. & von Heijne, G. (2001). Formation of helical hairpins during membrane protein integration into the endoplasmic reticulum membrane. Role of the N and C-terminal flanking regions. *J Mol Biol* **313**, 1171-9.
169. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A transmembrane helix dimer: Structure and implications. *Science* **276**, 131.
170. Getmanova, E., Patel, A. B., Klein-Seetharaman, J., Loewen, M. C., Reeves, P. J., Friedman, N., Sheves, M., Smith, S. O. & Khorana, H. G. (2004). NMR spectroscopy of phosphorylated wild-type rhodopsin: mobility of the phosphorylated C-terminus of rhodopsin in the dark and upon light activation. *Biochemistry* **43**, 1126-33.
171. Klein-Seetharaman, J., Yanamala, N. V., Javeed, F., Reeves, P. J., Getmanova, E. V., Loewen, M. C., Schwalbe, H. & Khorana, H. G. (2004). Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc Natl Acad Sci U S A* **101**, 3409-13.
172. Klein-Seetharaman, J., Reeves, P. J., Loewen, M. C., Getmanova, E. V., Chung, J., Schwalbe, H., Wright, P. E. & Khorana, H. G. (2002). Solution NMR spectroscopy of [alpha -15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc Natl Acad Sci U S A* **99**, 3452-7.
173. Schubert, M., Kolbe, M., Kessler, B., Oesterhelt, D. & Schmieder, P. (2002). Heteronuclear multidimensional NMR spectroscopy of solubilized membrane proteins: resonance assignment of native bacteriorhodopsin. *ChemBioChem* **3**, 1019-23.
174. Oxenoid, K., Kim, H. J., Jacob, J., Sonnichsen, F. D. & Sanders, C. R. (2004). NMR assignments for a helical 40 kDa membrane protein. *J Am Chem Soc* **126**, 5048-9.
175. Bennett, M., Yeagle, J. A., Maciejewski, M., Ocampo, J. & Yeagle, P. L. (2004). Stability of loops in the structure of lactose permease. *Biochemistry* **43**, 12829-37.
176. Katragadda, M., Alderfer, J. L. & Yeagle, P. L. (2001). Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin. *Biophys J* **81**, 1029-36.
177. Katragadda, M., Chopra, A., Bennett, M., Alderfer, J. L., Yeagle, P. L. & Albert, A. D. (2001). Structures of the transmembrane helices of the G-protein coupled receptor, rhodopsin. *J Pept Res* **58**, 79-89.
178. Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. (2002). NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207-11.
179. Rastogi, V. K. & Girvin, M. E. (1999). Structural changes linked to proton translocation by subunit c of the ATP synthase. *Nature* **402**, 263-8.
180. Howell, S. C., Mesleh, M. F. & Opella, S. J. (2005). NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury detoxification system. *Biochemistry* **44**, 5196-206.

181. Ma, D., Liu, Z., Li, L., Tang, P. & Xu, Y. (2005). Structure and dynamics of the second and third transmembrane domains of human glycine receptor. *Biochemistry* **44**, 8790-800.
182. Engelman, D. M. (2005). Membranes are more mosaic than fluid. *Nature* **438**, 578-80.
183. Fernandez, C., Hilty, C., Wider, G., Güntert, P. & Wüthrich, K. (2004). NMR structure of the integral membrane protein OmpX. *J Mol Biol* **336**, 1211-21.
184. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R. & Choe, S. (2005). NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* **307**, 1317-21.
185. Opella, S. J., Kim, Y. & McDonnell, P. (1994). Experimental nuclear magnetic resonance studies of membrane proteins. *Methods in enzymology* **239**, 536-560.
186. Prosser, R. S., Hwang, J. S. & Vold, R. R. (1998). Magnetically aligned phospholipid bilayers with positive ordering: a new model membrane system. *Biophys J.* **74**, 2405-18.
187. Poget, S. F. & Girvin, M. E. (2007). Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochim Biophys Acta* **1768**, 3098-106.
188. Zoonens, M., Catoire, L. J., Giusti, F. & Popot, J. L. (2005). NMR study of a membrane protein in detergent-free aqueous solution. *Proc Natl Acad Sci U S A* **102**, 8893-8.
189. Gohon, Y., Dahmane, T., Ruigrok, R. W., Schuck, P., Charvolin, D., Rappaport, F., Timmins, P., Engelman, D. M., Tribet, C., Popot, J. L. & Ebel, C. (2008). Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophys J* **94**, 3523-37.
190. Shi, Z., Peterson, R. W. & Wand, A. J. (2005). New reverse micelle surfactant systems optimized for high-resolution NMR spectroscopy of encapsulated proteins. *Langmuir* **21**, 10632-7.
191. Nath, A., Atkins, W. M. & Sligar, S. G. (2007). Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins. *Biochemistry* **46**, 2059-69.
192. Lyukmanova, E. N., Shenkarev, Z. O., Paramonov, A. S., Sobol, A. G., Ovchinnikova, T. V., Chupin, V. V., Kirpichnikov, M. P., Blommers, M. J. & Arseniev, A. S. (2008). Lipid-protein nanoscale bilayers: a versatile medium for NMR investigations of membrane proteins and membrane-active peptides. *J Am Chem Soc* **130**, 2140-1.
193. Cohen, L. S., Arshava, B., Estephan, R., Englander, J., Kim, H., Hauser, M., Zerbe, O., Ceruso, M., Becker, J. M. & Naider, F. (2008). Expression and biophysical analysis of two double-transmembrane domain-containing fragments from a yeast G protein-coupled receptor. *Biopolymers* **90**, 117-30.
194. Getz, E. B., Xiao, M., Chakrabarty, T., Cooke, R. & Selvin, P. R. (1999). A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal Biochem* **273**, 73-80.
195. Russ, W. P. & Engelman, D. M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **296**, 911-9.
196. Sal-Man, N., Gerber, D., Bloch, I. & Shai, Y. (2007). Specificity in transmembrane helix-helix interactions mediated by aromatic residues. *J Biol Chem* **282**, 19753-61.

197. Sanders, C. R. & Sonnichsen, F. (2006). Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44 Spec No**, S24-40.
198. Fernandez, C. & Wuthrich, K. (2003). NMR solution structure determination of membrane proteins reconstituted in detergent micelles. *FEBS Lett* **555**, 144-50.

Chapter 2

Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family

2.1 Introduction

Neuropeptide Y receptors, so-called Y receptors, are members of the rhodopsin-like G-protein coupled receptor (GPCR) family 1b. The neurohormones neuropeptide Y (NPY), peptide YY (PYY) and the pancreatic polypeptide (PP) target a heterologous population of at least five different receptor subtypes Y1, Y2, Y4, Y5 and y6[1]. Their physiological role in the regulation of blood pressure, memory retention, food uptake and seizure has been demonstrated. Y4 receptors for example have been shown to play a pivotal role in cardiac function, glucose metabolism in chronic pancreatitis patients, and mediation of intestinal absorption of electrolytes and water.[2] NPY and PYY posses a similar pharmacology displaying nanomolar affinities for all receptor subtypes[3], whereas PP binds with very high affinity and selectivity to the Y4 receptor.[4]

Little structural information is available for GPCRs. In fact, bovine rhodopsin for a long time was the only GPCR for which experimental coordinates at atomic resolution have been published [5] until very recently a structure for the b-adrenergic receptor appeared.[6; 7] The data of rhodopsin confirmed the arrangement of the 7 transmembrane (TM) bundle postulated based on the lower-resolution cryo-EM data [8], but also revealed the non-anticipated presence of a short anti-parallel beta-sheet in the N-terminal domain. In contrast, the N-terminal domain of the b-adrenergic receptor was shown to be disordered.[6]

The N-terminal domains of other GPCRs (sub)families are known to play important roles in ligand binding. All the hormone receptors from GPCR family 2 contain a conserved region in the N-terminal domain, which is responsible for ligand

binding.[9] The N termini from family 3 GPCRs are the largest among all GPCRs, comprising usually more than 500 amino acids.[9] Grafting and mutagenesis studies have demonstrated conserved serine and threonine residues in these domains are directly involved in ligand binding.[10] Surprisingly, the expressed N terminus alone can bind the ligand with affinity similar to the one from the full-length receptor.[11]

In contrast, the N-terminal domains from family 1 GPCRs have received little attention, most likely because of their short length, usually less than 70 amino acids. However, recent studies have suggested a pivotal role of N termini from GPCRs of this class in ligand recognition and binding[12; 13; 14]. Furthermore, mutagenesis data highlight the prominent role of charged residues for ligand binding[15; 16]. Koller demonstrated that the N terminus of the calcitonin-like receptor is not only essential for binding to the ligands but also presents a determinant for ligand specificity.[17] The 35 amino-terminal residues of CCR2, expressed as a membrane-bound fusion protein, bind to its ligand with an affinity similar to that of the intact, wild-type receptor, indicating that the N terminus is sufficient for ligand binding in that case.[18] Based on the mutagenesis data on the N terminus of CX3C receptor and previous studies, Chen has proposed a two-step binding model, which comprises ligand binding followed by receptor activation. Therein, the residues located in the N-terminal domain play distinct roles during the different processes[19].

Complementary to the biological work described above GPCR fragments have been also studied using NMR. For example, Pervushin investigated the N-terminal domain of bacteriorhodopsin, a protein that is structurally highly related to GPCRs, in SDS micelles[20], and Ulfers studied the extracellular domain of the neurokinin-1 receptor in DPC micelles[21]. Riek presented a high-quality 3D NMR structure of the extracellular domain of CRF-R2β in complex with the peptide antagonist astressin[22]. The group of Yeagle has determined conformational preferences for peptides corresponding to the cytosolic loops[23], the 6th TM helix[24] and the N-terminus[23] of rhodopsin and Pellegrini studied the cytosolic domain[25] and the extracellular loops[26; 27] of the PTH1 receptor in the presence of DPC micelles. Furthermore, we recently determined the conformation of a polypeptide corresponding to the 7th TM helix of the yeast Ste2p receptor extended by 40 residues from the cytosolic tail [28] when integrated into DPC micelles.

In this work we focus on structural studies of the isolated 41 residue N terminus of the Y4 receptor, a family 1b GPCR that is targeted by members of the NPY family.

The location of this segment in the context of the entire human Y4 receptor is shown in snake plot in Fig. S1(see Supp. Mat.). In addition, we investigate possible interactions with the hormones both qualitatively and quantitatively. By limiting the system of the study to just the N-terminal domain and with the help of various biophysical methods we were able to develop a rather detailed picture, that would presently be difficult to achieve using the entire receptor. Moreover, we report on the synthesis of the difficult to express N-terminal domain suggesting a generally useful method to produce these polypeptides in isotopically-labelled form. The structure of N-Y4 and its topology in the presence of DPC or SDS micelles was elucidated by high-resolution NMR techniques. While unstructured in solution, in the presence of micelles a hydrophobic segment associates with the micelle and folds into a a-helix. Chemical shift mapping revealed potential interaction sites between PP and N-Y4. SPR techniques quantified the strength of this interaction. Mutagenesis studies identified residues of PP that are likely to be important for binding N-Y4. The data indicate that the isolated N- Y4 is capable of weakly binding to PP, and that much of the binding affinity is due to electrostatic interactions. To simulate the receptor milieu the carboxyl terminus of N-Y4 was additionally conjugated to a C12 fatty amino alcohol (dodecylphospho-ethanolamine) chain thereby mimicking its conjugation to the first TM helix in the entire receptor. In this lipopeptide the structure of the N-Y4 was not significantly affected. The study shows that PP associates to the flexible, central segment of N-Y4 and we speculate that transient binding to the N-terminal domain may facilitate transferring PP from the membrane-bound state into the receptor binding pocket.

2.2 Results

2.2.1 Recombinant Production of N-Y4

The N terminus of the Y4 receptor comprises 41 residues and is highly water-soluble. However, attempts to express it in form of a soluble ubiquitin fusion in *E. coli* resulted in unspecific fragmentation. To circumvent this problem, the N-Y4 was expressed as a fusion to the highly insoluble protein ketosteroidisomerase (KSI), which resulted in accumulation of the fusion protein in inclusion bodies. A TEV protease cleavage site was introduced to facilitate removal of the fusion partner.[29; 30] The sequence recognized by the TEV protease is ENLYFQ with Q as the P1' residue. To achieve the natural peptide sequence after cleavage, the P1' residue was replaced with the first residue from the target sequence (here it is Met)[31], and an additional GSGSGS linker was inserted to prevent steric hindrance during cleavage.

A problem of the chosen strategy was that the fusion protein must be solubilized in detergent that is compatible with the active protease. After extensive detergent screening, we observed that the ionic detergent sarcosyl solubilizes the fusion protein while preserving TEV protease activity to some extent. As shown in Fig. 1 cleavage efficiency is around 40% allowing recovery of about 2mg of ^{15}N-labeled N-Y4 from 1L of culture.

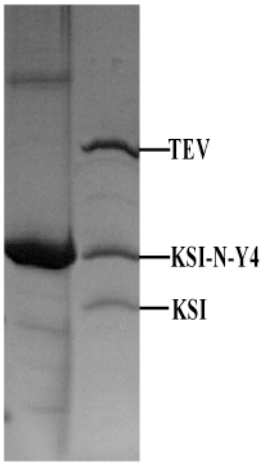


Figure 1: SDS-PAGE of the cleavage product of the ketosteroid isomerase- N-Y4 fusion after cleavage with the TEV protease. A size marker is shown on the left. Note that N-Y4 due to its small size cannot be detected on the gel.

2.2.2 The Structure of N-Y4

Although the size of the N-terminal domain is rather small, reduced chemical shift dispersion due to the fact that the peptide in water is largely unstructured complicated its analysis. Nevertheless, using 3D ^{15}N-resolved NOESY and TOCSY spectra it was possible to assign the ^{15}N,^{1}H-correlation map. Furthermore, no NOE crosspeaks between amide protons could be detected. Recording a second set of 2D and 3D spectra in the presence of DPC micelles resulted in large chemical shift changes in some parts of the sequence (see Fig.2).

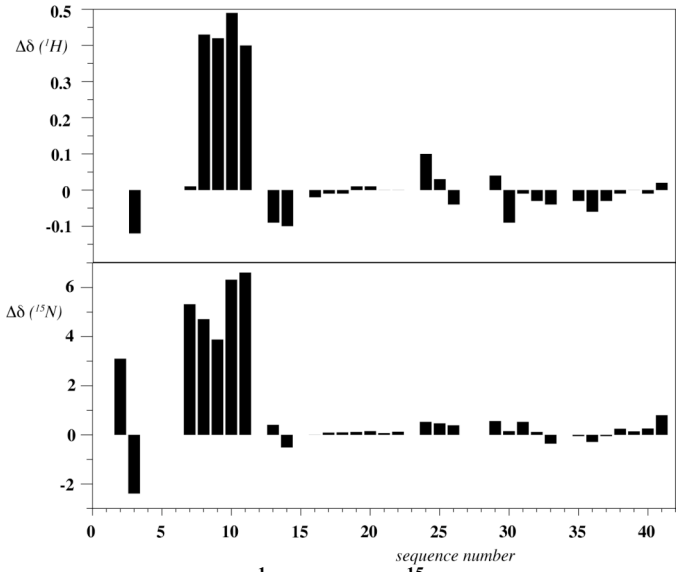


Figure 2: Differences of backbone amide ^{1}H (top) and ^{15}N (bottom) chemical shifts of N-Y4 in the presence and absence of DPC micelles.

Moreover, sequential NOEs between amide protons as well as HąHb (i,i+3) contacts usually only observed in helices were seen (see Supp. Mat.). A structure calculation using restraints derived from the NOESY spectra revealed the presence of a helical stretch encompassing residues 5 to 10 (shown in Fig.3).

To verify formation of stable secondary structure ^{15}N{^{1}H}-NOE spectra were recorded both in the absence as well as in the presence of DPC. The heteronuclear NOE sensibly reports on the rigidity of the backbone at the corresponding residue, with negative values characteristic of flexible parts and values larger than 0.5 usually observed in elements of secondary structure. The ^{15}N{^{1}H}-NOE data show dramatic differences in aqueous medium and DPC. Residues 1-27 have values <0 for N-Y4 in water whereas all of these residues have ^{15}N{^{1}H}-NOE values >0 in the DPC bound state (Fig. 4). Strikingly, residues 5-10 have a ^{15}N{^{1}H}-NOE >0.5. Interestingly, a

segment encompassing residues 26 to 33 is rather rigid, in both environments. We observed

Figure 3: Comparison of the structures calculated for N-Y4 in the presence of DPC (left) or SDS (right) micelles (only bonds from backbone atoms are depicted). Bonds from disordered residues 16-41 are not shown for clarity.

sequential amide proton contacts in that region for almost all residues, but the corresponding Hą Hb (i,i+3) contacts were generally missing. When comparing chemical shifts of amide protons in the two environments the largest differences were observed in that segment that obviously becomes structured in the presence of the micelle, indicating the presence of a nascent helix in that part. To conclude, the N terminus is largely unstructured in the absence of a membrane whereas a short helical stretch comprising a hydrophobic segment in the N terminus of the sequence is formed in presence of DPC micelles.

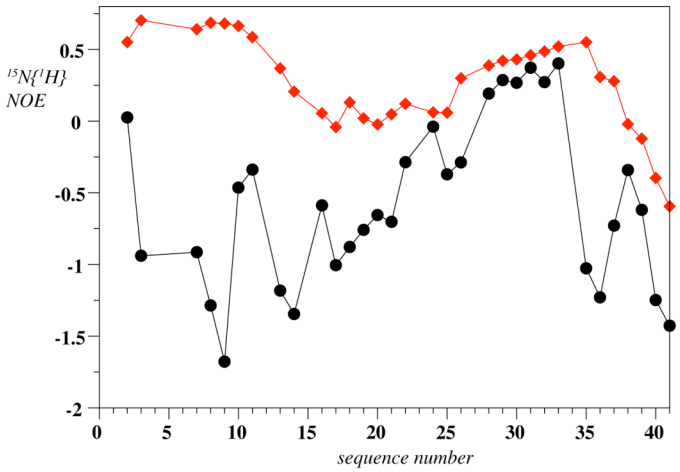


Figure 4: Values of the $^{15}N\{^{1}H\}$-NOE of N-Y4 in plain buffer (black spheres) and in the presence of DPC micelles (red diamonds). Data were recorded on 1mM samples at pH=5.6, 310K, at 700 MHz proton frequency.

Considering the importance of electrostatic interactions for ligand binding and to investigate whether (stabilizing) interactions of the N-terminal domain with the membrane head groups might be formed we further initiated structural studies of N-Y4 in the presence of SDS micelles, a negatively charged membrane mimetic. Values of the $^{15}N\{^{1}H\}$-NOE rapidly revealed that N-Y4 was not significantly better structured in this environment. Moreover, a structure calculation again revealed the presence of an a-helix spanning the region between residues 3 to 10. NOEs between sequential amide protons were seen at the C-terminal end from residue 36 on, but the corresponding Ha,Hb (i,i+3) contacts were missing, indicating that a transient helix is formed towards the C terminus. Interestingly, this part in the full-length receptor is connected to the first TM. In general, sequential amide proton contacts in the more flexible regions were stronger when compared to the data recorded in the presence of DPC suggesting that the negatively charged surface promotes the formation of transient helical structures to a slightly larger extent. This fact is particularly well-documented in the heteronuclear NOEs for residues of the segment encompassing residues 19-25, which is much less flexible in the presence of SDS micelles (see Fig. S8). But in general the structural features of the peptide in DPC and SDS were similar (for more data on the SDS-recorded sample see the Supp. Mat.)

2.2.3 Topology of Membrane-Association

The proximity of protons of the N-terminal domain to the micelle surface was probed by using micelle-integrating spin labels. The paramagnetic moiety of 5-doxyl stearic acid was shown to reside in the headgroup region[32] Consistent with the assumption that structuring of the N-terminal segment is induced by binding to the micelle, signals from the amide moieties within that segment experienced the largest signal reduction (see Fig. S9). The spin-label data indicate that the N-terminal helix is tightly associated with the micelle, whereas the central segment makes more transient contacts. Motions in that region are likely limited at both ends by the adjacent hydrophobic residues 24-30 and the membrane-anchored N-terminal helix. It was previously demonstrated that attenuations in helical regions of surface-associated peptides follow periodic patterns.[33; 34] The present data indicate that the helical region is not bound in a parallel fashion to the micelle-surface. Moreover, from the lack of a

clear pattern in the attenuation we conclude that this part is also not anchored in a precisely defined mode.

We have additionally tested whether binding of bPP to N-Y4 could possibly trigger dissociation of the N terminus from the micelle. However, no decrease of signal reduction from the spinlabel could be detected upon addition of bPP to micelle-bound N-Y4, indicating that N-Y4-micelle contacts are largely unchanged, even in the presence of a large excess of bPP (concentration ratio of N-Y4 to bPP 1:30) (data not shown). This indicates that bPP cannot initiate detachment of N-Y4 from the micelle surface, supporting the view that the contact site between bPP and Y-4 is not located in the helical segment of Y-4 and hence does not interfere with micelle association.

2.2.4 Immobilizing the N Terminus on the Membrane

In the native Y4 receptor the segment that has been studied in this work is connected to the first TM helix. In order to address whether anchoring of N-Y4 at its C-terminal end to the membrane influences the structure or the binding properties of the N-terminal domain a lipopeptide was chemically synthesized, in which receptor residues 1-41 were covalently linked at their C-terminus to dodecylethanolamine to provide stable anchoring of the lipopeptide in the micelles. The lipopeptide was prepared using standard amino-acid coupling chemistry, purified, and could be tightly integrated into the DPC micelles. A superposition of the NOESY spectra of N-Y4 and the lipopeptide in the presence of DPC micelles revealed that chemical shift differences are exclusively observed in vicinity of the lipid attachment site. Moreover, cross peaks between amide protons occur at identical positions, indicating that the secondary structure of both the peptides is highly similar. To conclude, anchoring of N-Y4 onto the micelle does not influence its secondary structure, which more likely is determined by partitioning of residues of the hydrophobic Leu-rich segment into the membrane. As evident from Fig. 3 the carboxyl terminal segment of N-Y4 possesses high flexibility both in the presence and in the absence of DPC micelles. Whether this will also be true when the C-terminus is linked to the first TM helix is presently under investigation.

2.2.5 Interaction between N-Y4 and Neuropeptides from the NPY Family

Possible interactions between peptides from the NPY family and N-Y4 were probed both by chemical shift mapping as well as by surface plasmon resonance (SPR). PP represents a natural ligand for the Y4 receptor, and accordingly the binding affinity between N-Y4 and PP was measured under physiological conditions (10mM HEPES pH 7.4, 150mM NaCl) both in absence and presence of DPC micelles. The data for chemical shift mapping were acquired using ^{15}N-labeled NPY,PP or PYY and unlabeled N-Y4 as well as using ^{15}N-labeled NY-4 and unlabelled neuropeptides. The shift mapping experiments revealed significant shift changes in the PP-N-Y4 interaction studies (see Fig. 5). Large changes in the PP/N-Y4 system occurred close to positions that were later on shown to be sensitive to replacement by Ala residues (vide infra). In addition, the shift changes involving PYY and NPY are generally much smaller compared to those with PP (data not shown).

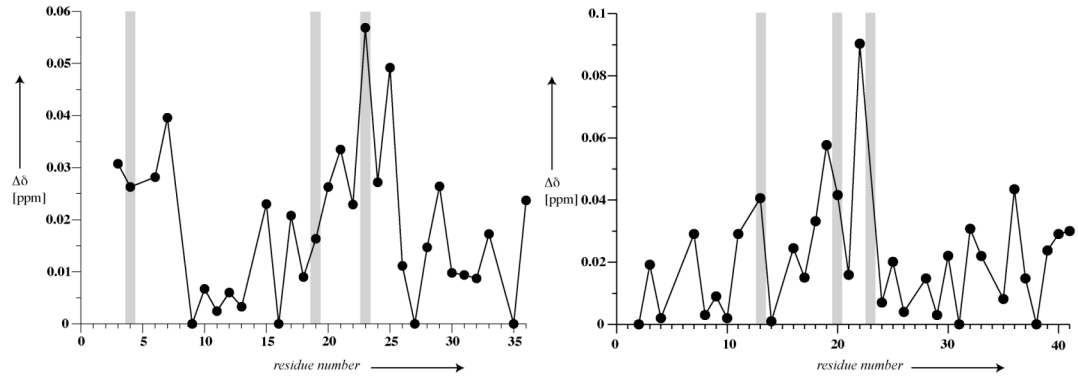


Figure 5: Differences of chemical shifts of amide proton and nitrogen frequencies of backbone resonances of bPP in the presence and absence of N-Y4 (Dd= d(bPP(N-Y4)) – d (bPP) (left) and of N-Y4 upon addition of bPP (right). Values are computed according to $Dd_c(^1H,^{15}N) = SQR [(Dd^1H)^2 + 0.2*(Dd^{15}N)^2]$. Positions at which mutations were performed (E4K, Q19R and E23A in PP and K13A, R20A and K23A in N-Y4) are indicated by grey bars.

The strength of the interaction of PP with N-Y4 was quantified by SPR in absence of detergent. Therein, the N-terminally biotinylated neuropeptides were immobilized on a Streptavidin-coated chip, and the cells were flushed with solutions of N-Y4 (see Fig. 6). The K_D derived from both kinetic and steady-state analysis was 50mM for bPP, whereas binding affinity for NPY and PYY was too low to be measured with this technique (> 1mM).

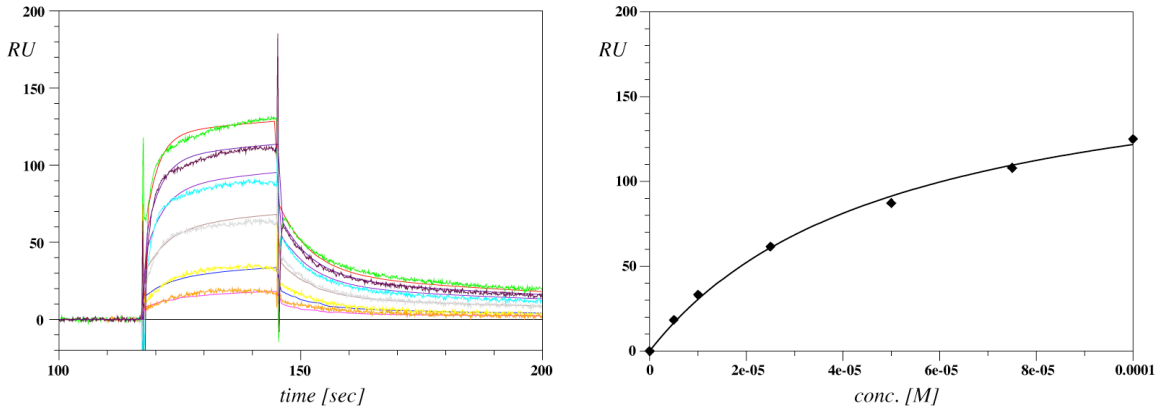


Figure 6: Left: SPR sensogram of the interaction of N-Y4 with bPP for various concentrations of N-Y4 (in the range of 5 to 100 µM). Right: Plot of the steady-state value of the sensograms *vs.* the concentration of N-Y4, used for extraction of the dissociation constant K_D.

Measuring binding of membrane-immobilized peptides towards N-Y4 by SPR methods is technically very challenging, and hence K_D in the presence of micelles were measured using NMR data by fitting changes in chemical shifts as derived from peak positions of the neuropeptides in [^{15}N,^{1}H]-HSQC spectra in the presence of varying amounts of N-Y4. For micelle-bound bPP the K_D to N-Y4 is approx. 600mM and experiments, in which varying amounts of PP were added to N-Y4, resulted in a very similar value. Apparently, the K_D in the presence of micelles is much lower than in the absence of micelles. This is not really surprising since it presents the affinity of the ligand towards the N-terminal domain in the presence of competing membrane binding, and hence reflects the difference in binding affinity between the two sites.

NPY and PYY possess 80% sequence identity between each other[35], while PP only shares about 50% homology to each of them. All these neuropeptides display a remarkable separation of charges along the sequence: The positively charged residues occur in the C-terminal half of PP from almost all organisms sequenced so far (see Table 1). In order to identify residues that may contribute significantly to the different pharmacological profiles of NPY/PYY and PP at the Y4 receptor we have aligned the sequences. Particular attention was paid to charged or aromatic residues that are known to be generally involved in GPCR-ligand interactions. The N termini of all Y receptor subtypes are generally negatively charged with the exception of N-Y4 that contains a net positive charge (see Table 1). Considering the high number of positive charges in N-Y4 and negative charges in the N-terminal half of bPP electrostatic interactions are likely to be responsible for binding, and such forces are also expected to result in the observed rather weak binding affinities.

Table 1. Sequence alignment of the principal members of the NPY family and of the N-terminal domains from the various Y receptor subtypes. Positions in bPP and hN-Y4 replaced by other amino acids in this work have been underlined.

pNPY: YPSKPDNPGE DAPAEDMARY YSALRHYINL ITRQRY-NH_2
pPYY: YPAKPEAPGE DASPEELSRY YASLRHYLNL VTRQRY-NH_2
bPP : APLEPEYPGD NATPEQMAQY AAELRRYINM LTRPRY-NH_2
* * * * * * * **** ****
hN-Y1: MNSTLFSQVE NHSVHSNFSE KNAQLLAFEN DDCHLPLAMI
hN-Y2: MGPIGAEADE NQTVEEMKVE QYGPQTTPRG ELVPDPEPEL IDSTKLIEVQ
hN-Y4: MNTSHLLALL LPKSPQGENR SKPLGTPYNF SEHCQDSVDV M
hN-Y5: MSFYSKQDYN MDLELDEYYN KTLATENNTA ATRNSDFPVW
DDYKSSVDDL Q

As depicted in Table 1 common acidic residues in PP, NPY and PYY are located at positions 6, 10 and 15. PP mutants E4K, Q19R and E23A were produced by site-directed mutagenesis in order to probe for the importance of differently charged residues between PP and NPY/PYY at these positions. The dissociation constant for Q19R-bPP was only marginally reduced to 89mM, whereas binding of E4K-bPP and E23A-bPP to N-Y4 was too weak to be detected by SPR. The data indicate that it is the additional negative charges in PP and their distribution along the sequence that may be important for its different binding affinities at the N-Y4.

In order to verify that electrostatic interactions between acidic residues of PP and basic residues in the N-Y4 are contributing to binding, the K13A, R20A and K22A mutants of the N-terminal domain of the Y4 receptor were synthesized and investigated by SPR. In all of these mutants binding to bPP was significantly reduced. The measured values for the K_D were 249 mM (R20A), 281mM (K22A) and for K13A binding was too weak to be detected by SPR. The combination of the mutagenesis studies performed on acidic residues of PP and basic residues of N-Y4 suggests that the binding affinity between the two is determined by electrostatic interactions to a large extent. In this work we have abstained from experiments in which residues in PP and N-Y4 were charged-reversed simultaneously because in those mutants electrostatics are likely to be perturbed in both molecules, and hence it is questionable whether activity could have been rescued.

2.3 Discussion

The mechanism for recognition of ligands by their receptors is of prime biological and pharmaceutical interest. Due to the enormous problems in expression, purification and reconstitution of sufficient amounts of GPCRs, little progress has been made in structural studies over the last decade, and so far bovine rhodopsin and the b-adrenergic receptor are the only GPCRs for which high-resolution X-ray data are published. In this work we have attempted to investigate the structure of the isolated N-terminal extracellular domain of the Y4 receptor, a GPCR targeted by hormones of the NPY family, and which binds to PP with very high affinity. Moreover, we determined the interaction with PP and the other members of the NPY family and investigated the role of specific residues for binding.

Structural studies of GPCR fragments could possibly suffer from the fact that interactions with the remainder of the receptor are missing that may be structurally relevant. As to the present analysis the N-terminal domain of the published crystal structure of b-adrenergic receptor was largely unstructured, and did not display interactions with other parts of this GPCR, in particular not with the extracellular loops. This supports our contention that the conformations of the N-terminal domains of a GPCR are not significantly determined by interactions with the remainder of the receptor. Such a study also allows us to directly define contributions of residues from the N-terminus of the Y-4 receptor to ligand binding.

While the N-terminal domain of Y4 is largely unfolded in solution upon binding to zwitterionic (DPC) or negatively charged (SDS) micelles, a hydrophobic segment comprising residues 5 to 10 forms a rather stable a-helix, and the nascent helix encompassing residues 26-35 is slightly rigidified. The central region and the C-terminal hexapeptide remain largely unstructured. The helical segment comprising residues 5 to 10 is entirely formed by hydrophobic residues. The structural data and the internal backbone dynamics of N-Y4 in the presence of zwitterionic (DPC) and anionic (SDS) headgroups display only minor differences indicating that the conformation does not depend on specific features of the surrounding lipids. Both formation of secondary structure and association with the membrane seem to be controlled by the hydrophobicity of the residues and their partitioning into the membrane[36] Strongly favorable values for the latter are encountered only in the a-

helical stretch and in the segment between residues 24 to 30, exactly those regions for which the spin-label data indicate proximity to the water-membrane interface. Spin-label, dynamics and structural data of Y-4 reveal the central segment to be rather flexible. The segregation of N-Y4 into structured and flexible regions is very similar in the presence of zitterionic or negatively charged lipid headgroups. As a consequence of these features it appears likely that this domain may perform larger movements on the membrane surface, and hence could possibly undergo various structural or translational transitions in order to interact with the extracellular loops or with the membrane-bound ligands. We like to mention at this point that the N-terminal domain of the b-adrenergic receptor was also disordered in the crystal structure from Kobilka[6; 7], and that the N-terminal domains from many other class-1 GPCRs are predicted to be largely unfolded. This indicates that the fact that N-Y4 is mainly flexible is likely not an artifact due to the usage of a receptor fragment but rather reflects a commonly encountered feature of these receptors.

We have recently proposed that binding of hormones from the NPY family to their receptors is preceded by association of the ligands to the membrane. According to ideas originally proposed by Kezdy and Kaiser[37; 38] and later developed into the membrane-compartment model by Schwyzer[39; 40] binding to the membrane reduces the search for the receptor to two dimensions, increases the concentration in the vicinity of the receptor and possibly induces conformations that facilitate receptor binding. Structural studies of porcine (p) NPY[33] and PYY[41] and of bovine (b) PP[42] bound to membrane-mimicking dodecylphosphocholine (DPC) micelles revealed large structural changes occurring during membrane association.[43] From this picture the important question arises how the hormones enter the binding pocket, once the membrane-bound species has laterally diffused along the membrane into the proximity of the receptor. The seven-helix bundle provides a rather rigid scaffold that does not allow large rearrangements of the extracellular loops in order to facilitate diffusion of the membrane-bound ligand into the binding pocket. Therefore the hormones need to detach from the membrane. Data for binding affinities of the hormones towards phospholipid membranes determined by us using SPR indicate that membrane binding is only moderate.[41] Any part of the receptor that possesses higher affinity to the peptides than the membrane does, and which could be accessed by a ligand that is in proximity to the membrane surface, may help to guide the ligand into the binding pocket. The N-terminal domains of the Y receptors are 40-50 amino acid

residue long polypeptide segments located in the extracellular space[44], and hence present potential interaction sites for the ligands. This work now indicates that at least for PP transient association with the N-Y4 may be part of the cascade of events leading to receptor activation. It should be emphasized here that transient binding to the N-terminal domain does not exclude larger structural changes in the conformations of loop residues that may occur later on when the ligands have diffused into the genuine receptor binding pockets. Such changes or rotations of the TM helices are believed to be important for receptor activation, and the above-described events merely serve to guide the ligand from the membrane-bound state into the binding pocket.

Binding of PP to the N-terminal domain of the Y4 receptor, which is often referred to as the PP-preferring receptor, is moderate with a dissociation constant of about 50mM. NPY and PYY, two hormones from the NPY family with very similar pharmacology and high sequence similarity with respect to each other, do not bind to this domain. Sequence alignments reveal that PP overall is more negatively charged than NPY or PYY, particularly in the N-terminal region, and our studies show that replacement of E4 or E23 in PP largely abolished binding to N-Y4. Furthermore, introduction of Arg into position 19 lead to only marginal changes in binding affinity. The N-Y4 domain, in contrast to the N-terminal domains from all other receptor subtypes, contains a comparably large number of positively charged residues (K13, R20 and K22), which are also relatively close to each other in sequence. Their replacement by Ala as described above leads to significant losses in binding affinity. To conclude taking the importance of acidic PP and basic N-Y4 residues into account we speculate that electrostatic interactions between PP and N-Y4 are crucial for this interaction. However, it must be emphasized that *a priori* it is not clear in our case whether residues from the N terminus are interacting with residues from the extracellular loops thereby modulating the effective charge experienced by the peptides. This question can only be addressed experimentally with confidence when structural studies of the full-length receptor in a functional state become available.

Unfortunately, not much pharmacological data is available for the entire Y4 receptor. In case of the human Y1 receptor an Asp residue at the interface between TM helix 6 and the third extracellular loop was proposed to contribute largely to binding NPY[45] in the full-length Y1 receptor. Considering that Asp at this position is conserved amongst all Y receptor subtypes it was speculated that this residue

generally contributes to binding in all subtypes. Nicole et al. investigated the role of this Asp$^{6.59}$ in more detail[46] and verified the proposed interaction of Arg33 or Arg35 with acidic third extracellular loop (ECL3) residues in the other Y receptor subtypes. Our data now indicate that in addition to the above-described interaction additional contacts between acidic residues of PP and basic residues of the N-terminal domain of the Y4 receptor may contribute to binding. Association of the N-Y4 with PP may be therefore not only be of transient nature helping the ligand to be transferred from the membrane-bound state into the receptor binding pocket, but may also exist in the ligand-bound state, contributing to the high binding affinity and selectivity of PP at the Y4 receptor.

2.4 Conclusions

Based on the data described above, we speculate that the N-terminal domain of the Y4 receptor may help in transferring PP from the membrane-bound state into the receptor binding pocket (see Fig. 7). As proposed by us in case of ligands of the Y receptors[43] PP initially associates with the membrane. By binding to the membrane the effective concentration in vicinity of the receptor is increased, the search is reduced from three to two dimensions, and conformations closer to those of the bound state may be induced according to the membrane-compartment model.[39; 40] BIACore data of PP binding to phospholipid surfaces indicated that binding to membranes is moderate.[47] Accordingly, an equilibrium is formed, in which PP rapidly diffuses on and off the membrane, but mostly remains in vicinity of the membrane. When PP has diffused into proximity of the receptor where interactions with the latter can occur it may transiently bind to N-Y4 from solution. Whether the complex of PP and N-Y4 itself will move into vicinity of the extracellular loops, or whether the position of N-Y4 is fixed by interactions with the membrane or the remaining portion of the receptor is presently unclear.

A scenario, in which N-Y4-bound PP would be transferred into the binding pocket by a translational movement of parts of the N-terminal domain is at least compatible with the experimental data. These indicate that the binding region for PP is located in its central segment, which at the same time is the only part of N-Y4 that is not making significant contacts with the membrane surface, and which also possesses sufficient internal flexibility to allow the necessary movements. We presently favor a view that describes the N-terminal domain as a large flexible loop, anchored onto the membrane at the amino terminus via the membrane-associated helix and at the C terminus via the first TM. This view is also supported by the recent crystal structures of the b-adrenergic receptor in which the N-terminal domain is so flexible that electron density in this part could not be traced.[6; 7] We have now initiated work on constructs that include parts of the TM bundle to see whether conformational preferences of N-Y4 are influenced by the remainder of the receptor.

2.5 Materials and Methods

Expression of the N-Y4 sequence as a soluble fusion to ubiquitin resulted in heterogeneous fragmentation. In order to prevent *in-vivo* processing the N-terminal domain was fused to the highly insoluble protein ketosteroidisomerase that is encoded in the commercial plasmid pET 31b, from which it was liberated by cleavage with the TEV protease in mild detergent.

2.5.1 Plasmid Construction, Expression and Purification of N-Y4

The cDNA of the Y4 receptor was obtained from the University of Missouri-Rolla (UMR) cDNA Resource Center. The following two primers were used to amplify the cDNA corresponding to N-Y4 by PCR. Forward primer: GCG*CTCGAG***GGTTCCGGTTCCGGTTCC**GAAAACCTGTACTTCCAGATGA ACACCTCTCACCTGCTGGC, in which italic letters denote a XhoI cleavage site, bold letters denote a Gly-Ser linker sequence and underlined letters identify a TEV cleavage sequence; backward primer: CTG*GCTGAGC*TCACATCACGTCCACGGAATCCT with italic letters denoting an EspI cleavage site. The amplified PCR product and the target vector, pET 31b (Novagen), were simultaneously digested with XhoI and EspI, and ligated into the vector with T4 ligase. The construct was confirmed by DNA sequencing (Synergene Biotech, Switzerland). All mutants were constructed by site-directed mutagenesis using the QuikChange Kit (Stratagene, USA).

The fusion protein was expressed in inclusion bodies using the BL21(DE3) *E.coli* strain. Protein expression was performed by growing cells at 37°C using minimal media containing ^{15}N-NH_4Cl as the sole nitrogen source for ^{15}N labeled peptide. 1mM IPTG was added to induce protein expression when the OD_{600} reached 0.8 and cells were harvested after 5-6 hours. The fusion protein was purified from inclusion bodies in 6M guanidinium hydrochloride by Ni-NTA affinity chromatography. After removal of GdnHCl by dialysis the precipitated fusion protein was solubilized in 50mM Tris pH 8.0 in the presence of 2% N-lauryl sarcosine upon sonication to a final concentration of 2mg/ml. The resulting solution was dialyzed against a 20-fold excess of 50 mM Tris, pH 8.0 for 4-6 times. The solution was diluted 10 times with 50 mM

Tris pH 8.0 and EDTA and DTT were added to a final concentration of 0.5 mM and 1 mM, respectively. TEV protease was added to a final concentration of 100 mM and the cleavage mixture was kept at 4°C over night. The target peptide was purified by C18-RP-HPLC (Vydac, USA) and the correctness of the peptide was verified by MALDI-TOF MS: ^{15}N labeled N-Y4: 4614 Da (theoretical mass (for 100% labeling): 4611.1 Da).

2.5.2 Synthesis and Purification of the Neuropeptides and of Unlabelled N-terminal Fragments

^{15}N-labeled peptides from the NPY family were expressed as soluble fusions to Ubiquitin. Ubiquitin was liberated from the neuropeptide using the yeast ubiquitin hydrolase, and C-terminal amidation was performed using the a-amidating peptidyl glycine amidase (PAM). We have used the protocols for expression, ubiquitin cleavage and C-terminal amidation many times before and described in them much detail elsewhere, e.g. in Bader et al.[33]

Wild-type and mutant N-Y4 peptides and peptides from the NPY family containing ^{15}N nuclei at natural abundance were prepared by solid-phase peptide synthesis using a robot system (ABI433A, Applied Biosystems). 2-chlorotrityl chloride resin preloaded with Fmoc-Met-OH was used to assemble the linear peptide using standard Fmoc chemistry (20% piperidine in DMF for Fmoc deprotection, 4 equiv. HOBt/HBTU for activation, diisopropylethylamine as base, and *N*-methylpyrrolidone as solvent). The peptides were cleaved from the resin and deprotected with TFA/water/1,2-ethanedithiol/triisopropylsilane 95/2.5/2.5/2.5. The product was lyophilized and purified by C18 RP-HPLC and correctness was confirmed by ESI-MS: wild-type N-Y4: 4556.8 Da (theoretical mass: 4556.1 Da); K13A N-Y4: 4501 Da (theoretical mass: 4499 Da); R20A N-Y4: 4473 Da (theoretical mass: 4471 Da); K22A N-Y4: 4501 Da (theoretical mass: 4499 Da).

In order to synthesize the N-terminally biotinylated forms the peptides were mixed with biotin-$(PEO)_4$-NHS-propionate (Molecular Biosciences, USA) in a 1:2 ratio in 100mM phosphate buffer, pH 7.0 and incubated for 2 hours at RT and afterwards purified by C18 RP-HPLC and confirmed by ESI-MS. To confirm that in case of E4K-bPP, the biotin was coupled to the N-terminus instead of the side chain of lysine, the biotinylated peptide was first digested with pepsin, and subsequently the

fragment containing residue 1-16 was analyzed by MALDI-TOF MS-MS. The result from this analysis demonstrated that the biotin was exclusively coupled to the N terminus.

2.5.3 Dodecylphosphoethanolamine Coupling to the Carboxyl Terminus of N-Y4

The peptide from solid-phase peptide synthesis was cleaved off the resin with TFA (0.8 vol %) in DCM with all the protecting groups remaining intact. Following removal of solvents the protected peptide was precipitated in cold water, lyophilized and redissolved in DMF. The solution was stirred at RT for 5 hours with 3 equivalents of dodecylphosphoethanolamine (3 equiv.) in presence of of HATU (1 equiv.), HOAt (1 equiv.) and of DIEA(1.5 equiv.). After extraction with a ethyl-acetate:water mixture (1:1 v/v) the lipopeptide was deprotected under the same conditions as described above. Finally, the lipopeptide was purified by C4 RP-HPLC (Vydac, USA), lyophilized and purity higher than 95% was confirmed by MALDI-TOF-MS: 4848 Da (theoretical mass: 4847.1 Da) and LC-MS.

2.5.4 NMR Experiments

All samples of N-Y4 for structural studies were measured at 1mM concentration, 40mM d-MES at pH 5.6. For measurements mimicking membrane environments 300mM d_{38}-DPC or 300mM d_{25}-SDS were added. All experiments were performed at 700 MHz, 310K using a triple-resonance cryoprobe. Resonance assignments were initially performed in the absence of DPC or SDS using [^{15}N,^{1}H]-HSQC, 3D [^{15}N,^{1}H]-HSQC-TOCSY (80ms mixing time) and 300ms 3D [^{15}N,^{1}H]-HSQC-NOESY experiments. Details of the spectroscopy were similar to those described by us earlier.[48] Spectra were analyzed using the programs CARA[49] and XEASY.[50] After nearly complete resonance assignments in water were obtained, a 200ms 3D [^{15}N,^{1}H]-HSQC-NOESY was recorded in the presence of DPC, and the assignments in water adjusted to the DPC spectra. Upper distance restraints in DPC or SDS were then derived from 50ms 2D NOESY spectra. Internal backbone dynamics were studied by measuring a ^{1}H-detected version of a ^{15}N{^{1}H}-NOE experiment. Structures were computed based on upper-distance restraints derived from the NOESY spectra using the program CYANA[51; 52] following the standard simulated annealing protocol.

$^{15}N\{^{1}H\}$-NOEs were computed from the ratio of integrals from signals in the presence to those in the absence of amide proton irradiation.[53] Chemical shifts of the $^{15}N,^{1}H$-correlation map in the absence and full assignments in the presence of DPC and SDS can be found in the Supp. Mat. Proton chemical shifts were referenced to the water line, taken as 4.63 ppm at 310K, from which the nitrogen scale was derived indirectly through multiplication with the factor $g(^{15}N)/g(^{1}H)$.

The coordinates, chemical shift values and heteronucelar NOEs of NY-4 in the presence of SDS and DPC have been deposited in the BMRB database under the accession number 15708.

2.5.5 Membrane-Association Topology Using Spin Labels

In the spin label studies $[^{15}N,^{1}H]$-HSQC spectra of 0.5mM solutions of ^{15}N-N-Y4 containing 300mM DPC were measured in absence and presence of 7mM and 8.8mM 5-doxyl and 16-doxyl stearic acid, respectively. Signal attenuation was computed from the ratio of integrals from peaks in the corresponding spectra. The signal attenuation in the presence of the spin label is related to proximity of protons to the label. In another set of experiments 0.1mM ^{15}N-labeled N-Y4 was mixed with various concentrations of bPP in order to test whether N-Y4 is released from the micelle upon interaction with PP.

2.5.6 Surface Plasmon Resonance (SPR) Studies

HBS buffer (10mM HEPES, pH 7.4, 150mM NaCl, 3.4mM EDTA, 0.005% P20) was used as the running buffer to achieve physiological pH. N-terminally biotinylated neuropeptides were immobilized onto the sensor chip SA (BiaCore, Sweden), which contains a streptavidin-coated surface, resulting in about 200 response units (RU) on a BIAcore 1000 instrument (BIAcore, Sweden). Different concentrations of N-Y4 spanning a range of 5 to 100μM were applied to the surface for 30 seconds at a flow-rate of 20ml/min at 25°C. After each injection of analytes, the flow-cell was flushed with regeneration buffer (1M NaCl, 50mM NaOH) for 30 seconds. Since unspecific binding at concentrations higher than 100μM occurred, K_D larger than 100μM could not be determined precisely. Nevertheless, trends in reduction of binding could still be

computed from a limited set of data points, in which values at high concentrations were excluded from the analysis. All sensograms were analyzed with the BIA evaluation software using a two-state binding model.

2.6 Acknowledgments

We like to deeply acknowledge valuable discussion with F. Naider and R. Bader. We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173)

2.7 References

1. Michel, M. C., Beck-Sickinger, A., Cox, H., Doods, H. N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T. & Westfall, T. (1998). XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors. *Pharmacol Rev* **50**, 143-50.
2. Balasubramaniam, A. (2002). Clinical potentials of neuropeptide Y family of hormones. *Am J Surg* **183**, 430-4.
3. Larhammar, D., Wraith, A., Berglund, M. M., Holmberg, S. K. & Lundell, I. (2001). Origins of the many NPY-family receptors in mammals. *Peptides* **22**, 295-307.
4. Berglund, M. M., Lundell, I., Eriksson, H., Soll, R., Beck-Sickinger, A. G. & Larhammar, D. (2001). Studies of the human, rat, and guinea pig Y4 receptors using neuropeptide Y analogues and two distinct radioligands. *Peptides* **22**, 351-6.
5. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-45.
6. Rasmussen, S. G. F., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., Schertler, G. F. X., Weis, W. I. & Kobilka, B. K. (2007). Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor. *Nature* **450**, 383-U4.
7. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-65.
8. Schertler, G. F., Villa, C. & Henderson, R. (1993). Projection structure of rhodopsin. *Nature* **362**, 770-2.
9. Harmar, A. J. (2001). Family-B G-protein-coupled receptors. *Genome Biol* **2**, REVIEWS3013.
10. O'Hara, P. J., Sheppard, P. O., Thogersen, H., Venezia, D., Haldeman, B. A., McGrane, V., Houamed, K. M., Thomsen, C., Gilbert, T. L. & Mulvihill, E. R. (1993). The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41-52.
11. Okamoto, T., Sekiyama, N., Otsu, M., Shimada, Y., Sato, A., Nakanishi, S. & Jingami, H. (1998). Expression and purification of the extracellular ligand binding region of metabotropic glutamate receptor subtype 1. *J Biol Chem* **273**, 13089-96.
12. DeMartino, J. A., Van Riper, G., Siciliano, S. J., Molineaux, C. J., Konteatis, Z. D., Rosen, H. & Springer, M. S. (1994). The amino terminus of the human C5a receptor is required for high affinity C5a binding and for receptor activation by C5a but not C5a analogs. *J Biol Chem* **269**, 14446-50.
13. Dettin, M., Zanchetta, M., Pasquato, A., Borrello, M., Piatier-Tonneau, D., Di Bello, C. & De Rossi, A. (2003). CCR5 N-terminus peptides enhance X4 HIV-1 infection by CXCR4 up-regulation. *Biochem Biophys Res Commun* **307**, 640-6.

14. Ho, H. H., Du, D. & Gershengorn, M. C. (1999). The N terminus of Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor is necessary for high affinity chemokine binding but not for constitutive activity. *J Biol Chem* **274**, 31327-32.
15. Blanpain, C., Doranz, B. J., Vakili, J., Rucker, J., Govaerts, C., Baik, S. S., Lorthioir, O., Migeotte, I., Libert, F., Baleux, F., Vassart, G., Doms, R. W. & Parmentier, M. (1999). Multiple charged and aromatic residues in CCR5 amino-terminal domain are involved in high affinity binding of both chemokines and HIV-1 Env protein. *J Biol Chem* **274**, 34719-27.
16. Hawtin, S. R., Wesley, V. J., Simms, J., Argent, C. C., Latif, K. & Wheatley, M. (2005). The N-terminal juxtamembrane segment of the V1a vasopressin receptor provides two independent epitopes required for high-affinity agonist binding and signaling. *Mol Endocrinol* **19**, 2871-81.
17. Koller, D., Born, W., Leuthauser, K., Fluhmann, B., McKinney, R. A., Fischer, J. A. & Muff, R. (2002). The extreme N-terminus of the calcitonin-like receptor contributes to the selective interaction with adrenomedullin or calcitonin gene-related peptide. *FEBS Lett* **531**, 464-8.
18. Monteclaro, F. S. & Charo, I. F. (1997). The amino-terminal domain of CCR2 is both necessary and sufficient for high affinity binding of monocyte chemoattractant protein 1. Receptor activation by a pseudo-tethered ligand. *J Biol Chem* **272**, 23186-90.
19. Chen, Y., Green, S. R., Almazan, F. & Quehenberger, O. (2006). The amino terminus and the third extracellular loop of CX3CR1 contain determinants critical for distinct receptor functions. *Mol Pharmacol* **69**, 857-65.
20. Pervushin, K. V., Orekhov, V., Popov, A. I., Musina, L. & Arseniev, A. S. (1994). Three-dimensional structure of (1-71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by ^{15}N-^{1}H heteronuclear NMR spectroscopy. *Eur. J Biochem* **219**, 571-83.
21. Ulfers, A. L., Piserchio, A. & Mierke, D. F. (2002). Extracellular domains of the neurokinin-1 receptor: structural characterization and interactions with substance P. *Biopolymers* **66**, 339-49.
22. Grace, C. R., Perrin, M. H., Gulyas, J., Digruccio, M. R., Cantle, J. P., Rivier, J. E., Vale, W. W. & Riek, R. (2007). Structure of the N-terminal domain of a type B1 G protein-coupled receptor in complex with a peptide ligand. *Proc Natl Acad Sci U S A* **104**, 4858-63.
23. Yeagle, P. L., Salloum, A., Chopra, A., Bhawsar, N., Ali, L., Kuzmanovski, G., Alderfer, J. L. & Albert, A. D. (2000). Structures of the intradiskal loops and amino terminus of the G-protein receptor, rhodopsin. *J Pept Res* **55**, 455-65.
24. Chopra, A., Yeagle, P. L., Alderfer, J. A. & Albert, A. D. (2000). Solution structure of the sixth transmembrane helix of the G-protein-coupled receptor, rhodopsin. *Biochim Biophys Acta* **1463**, 1-5.
25. Pellegrini, M., Royo, M., Chorev, M. & Mierke, D. F. (1996). Conformational characterization of a peptide mimetic of the third cytoplasmic loop of the G-protein coupled parathyroid hormone/parathyroid hormone related protein receptor. *Biopolymers* **40**, 653-66.
26. Pellegrini, M. & Mierke, D. F. (1999). Structural characterization of peptide hormone/receptor interactions by NMR spectroscopy. *Biopolymers* **51**, 208-20.

27. Mierke, D. F., Mao, L., Pellegrini, M., Piserchio, A., Plati, J. & Tsomaia, N. (2007). Structural characterization of the parathyroid hormone receptor domains determinant for ligand binding. *Biochem Soc Trans* **35**, 721-3.
28. Neumoin, A., Arshava, B., Becker, J., Zerbe, O. & Naider, F. (2007). NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor from Saccharomyces cerevisiae. *Biophys J* **93**, 467-82.
29. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993-1000.
30. van den Berg, S., Lofdahl, P. A., Hard, T. & Berglund, H. (2006). Improved solubility of TEV protease by directed evolution. *J Biotechnol* **121**, 291-8.
31. Kapust, R. B., Tozser, J., Copeland, T. D. & Waugh, D. S. (2002). The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* **294**, 949-55.
32. Papavoine, C. H., Aelen, J. M., Konings, R. N., Hilbers, C. W. & Van de Ven, F. J. (1995). NMR studies of the major coat protein of bacteriophage M13. Structural information of gVIIIp in dodecylphosphocholine micelles. *Eur J Biochem* **232**, 490-500.
33. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. (2001).Structure and dynamics of micelle-bound neuropeptide Y: compasiron with unligated NPY and implications for receptor selection. *J Mol Biol* **305**, 307-392.
34. Respondek, M., Madl, T., Gobl, C., Golser, R. & Zangger, K. (2007). Mapping the orientation of helices in micelle-bound peptides by paramagnetic relaxation waves. *J Am Chem Soc* **129**, 5228-34.
35. Conlon, J. M. (2002). The origin and evolution of peptide YY (PYY) and pancreatic polypeptide (PP). *Peptides* **23**, 269-78.
36. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109-24.
37. Kaiser, E. T. & Kezdy, F. J. (1983). Secondary structures of proteins and peptides in amphiphilic environments. *Proc Natl Acad Sci U S A* **80**, 1137-43.
38. Kaiser, E. T. & Kezdy, F. J. (1984). Amphiphilic secondary structure: design of peptide hormones. *Science* **223**, 249-55.
39. Sargent, D. F. & Schwyzer, R. (1986). Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc Natl Acad Sci U S A* **83**, 5774-8.
40. Schwyzer, R. (1994). 100 Years Lock-and-Key Concept: Are Petide Keys Shaped and Guided to Their Receptors by the Target Cell Membrane? *Biopolymers* **37**, 5-16.
41. Lerch, M., Mayrhofer, M. & Zerbe, O. (2004). Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J Mol Biol* **339**, 1153-68.
42. Lerch, M., Gafner, V., Bader, R., Christen, B., Folkers, G. & Zerbe, O. (2002). Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles. *J Mol Biol* **322**, 1117-33.
43. Bader, R. & Zerbe, O. (2005). Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state? *ChemBioChem* **6**, 1520-34.

44. Larhammar, D. (1996). Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **62**, 1-11.
45. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. (1994). Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding. *J Biol Chem* **269**, 2863-9.
46. Merten, N., Lindner, D., Rabe, N., Rompler, H., Morl, K., Schoneberg, T. & Beck-Sickinger, A. G. (2007). Receptor subtype-specific docking of Asp6.59 with C-terminal arginine residues in Y receptor ligands. *J Biol Chem* **282**, 7543-51.
47. Lerch, M., Kamimori, H., Folkers, G., Aguilar, M.-I., Beck-Sickinger, A. G. & Zerbe, O. (2005). Strongly Altered Receptor Binding Properties in PP and NPY Chimera are Accompanied by Changes in Structure and Membrane Binding. *Biochemistry* **44**, 9255 - 9264.
48. Bader, R. (2001). Structural Aspects of Neuropeptide Y: Implications of the membrane-bound State for Receptor Recognition and Subtype Selectivity Studied by NMR. No. 14129, ETH Zurich.
49. Keller, R. (2004). *The Computer Aided Resonance Assignment*, CANTINA Verlag, Goldau.
50. Bartels, C., Xia, T.-h., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The program XEASY for computer-supported spectral analysis of biological macromolecules. *J Biomol NMR* **6**, 1-10.
51. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78.
52. Guntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283-98.
53. Noggle, J. H. & Schirmer, R. E. (1971). *The Nuclear Overhauser Effect - Chemical Applications*, Academic Press, New York.

2.8 Supplementary Materials

Fig. S1: "Snake"-plot type presentation of the human Y4 receptor. The plot was downloaded from the GPCR.org website. Note that to enhance clarity not all residues from the N-terminal domain and the long loops are shown

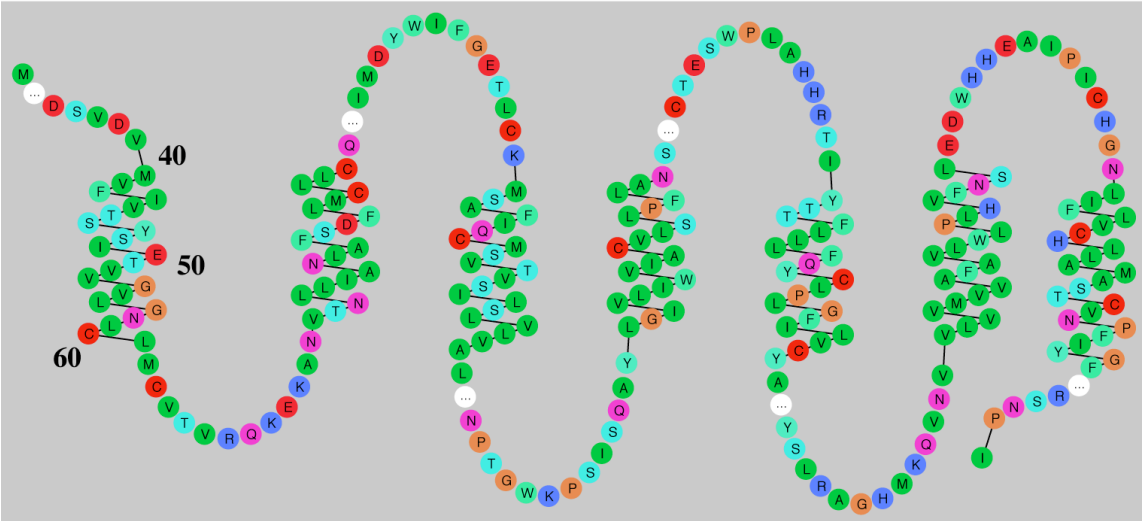


Fig. S2: Summary of the meaningful distance restraints as derived from the unambiguously assigned inter-residue NOEs between backbone H^N,H^α and H^β of N-Y4 bound to DPC micelles.

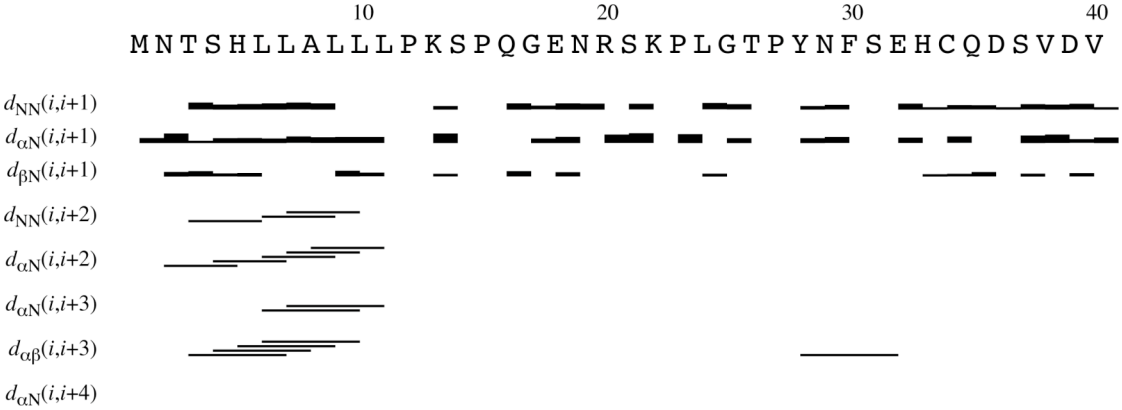


Fig. S3: Expansion of the 100ms NOESY displaying the region involving NOEs between sequential amide protons (DPC micelles)

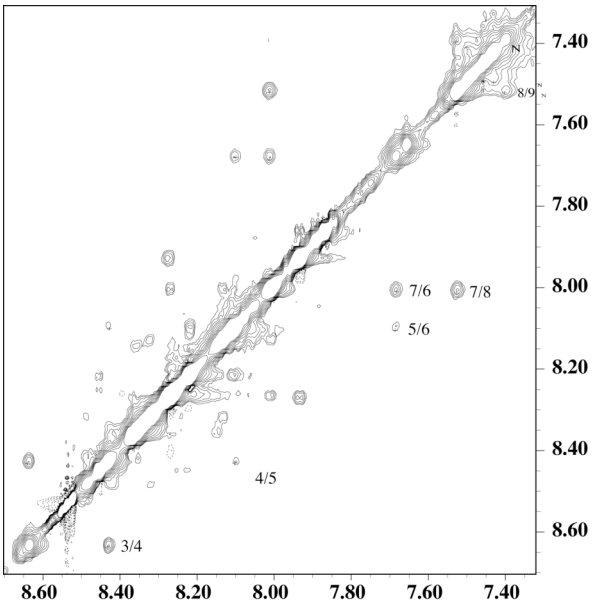


Table S4: Chemical shifts for N-Y4 bound to DPC micelles [a] referenced to the signal of residual HDO at 4.63 ppm

Residue	H^N	H^{α} [a]	H^{β} [b]	others
Met 1	-	4.07	-, -	g^{CH_2} -, -; e^{CH_3} -
Asn 2	8.99	4.92	2.92, 2.80	d^{NH_2} 7.66, 6.90
Thr 3	8.63	3.89	4.20	g^{CH_3} 1.20; gOH -
Ser 4	8.43	4.14	3.91, 3.91	gOH -
His 5	8.11	4.50	3.25, 3.25	d^1NH -; d^2H 6.89; e^1H 7.66; e^2NH -
Leu 6	7.69	3.98	1.74, 1.74	gH 1.67; d^{CH_3} 0.90, 0.84
Leu 7	8.01	3.91	1.75, 1.75	gH 1.58; d^{CH_3} 0.85, 0.85
Ala 8	7.52	4.03	1.44	
Leu 9	7.40	4.09	1.89, 1.89	gH 1.47; d^{CH_3} 0.83, 0.83
Leu 10	7.43	4.20	1.72, 1.72	gH 1.55; d^{CH_3} 0.83, 0.79
Leu 11	7.50	4.51	1.65, 1.65	gH 1.50; d^{CH_3} 0.87, 0.87
Pro 12		-	2.29, 1.86	g^{CH_2} 1.98, 1.98; d^{CH_2} 3.54, 3.69
Lys 13	8.35	4.30	1.70, 1.70	g^{CH_2} 1.42, 1.42; d^{CH_2} 1.78, 1.78; e^{CH_2} 2.96, 2.96; z^{NH_3} -
Ser 14	8.30	4.68	3.81, 3.81	gOH -
Pro 15		4.40	2.24, 1.88	g^{CH_2} 1.97, 1.97; d^{CH_2} 3.69, 3.79
Gln 16	8.37	4.27	2.08, 1.94	g^{CH_2} 2.33, 2.33; e^{NH_2} 7.45, 6.78
Gly 17	8.27	3.90, 3.90		
Glu 18	8.22	4.26	2.01, 1.88	g^{CH_2} 2.24, 2.24; eH -
Asn 19	8.45	4.64	2.79, 2.71	d^{NH_2} 7.53, 6.86
Arg 20	8.26	4.31	1.85, 1.71	g^{CH_2} 1.58, 1.58; d^{CH_2} 3.14, 3.14; eNH -; h^{NH_2} -, -
Ser 21	8.22	4.37	3.80, 3.80	gOH -
Lys 22	8.10	4.57	1.65, 1.65	g^{CH_2} 1.76, 1.76; d^{CH_2} 1.41, 1.41; e^{CH_2} 2.95, 2.95; z^{NH_3} -
Pro 23		4.40	2.27, 2.27	g^{CH_2} 1.98, 1.98; d^{CH_2} 3.55, 3.55
Leu 24	8.34	4.26	1.63, 1.63	gH 1.55; d^{CH_3} 0.86, 0.86
Gly 25	8.27	3.92, 3.92		
Thr 26	7.94	4.53	4.08	g^{CH_3} 1.15; gOH -
Pro 27		-	2.47, 2.14	g^{CH_2} 1.96, 1.82; d^{CH_2} 3.75, 3.57
Tyr 28	7.91	4.34	2.79, 2.79	dH 6.92, 6.92; eH 6.71, 6.71; hOH -
Asn 29	8.08	4.61	2.71, 2.59	d^{NH_2} 7.48, 6.79
Phe 30	8.21	4.44	3.13, 3.05	dH 7.24, 7.24; eH 7.14, 7.14; zH -
Ser 31	8.07	4.39	3.81, 3.81	gOH -
Glu 32	8.08	4.16	1.94, 1.85	g^{CH_2} 2.21, 2.21; eH -
His 33	8.24	4.59	3.28, 3.09	d^1NH -; d^2H -; e^1H -; □NH -
Cys 34	8.20	4.39	2.86, 2.86	□SH -
Gln 35	8.43	4.25	2.08, 1.96	□CH□ 2.31, 2.31; □NH□ 7.43, 6.76
Asp 36	8.25	4.64	2.76, 2.64	□H -
Ser 37	8.13	4.42	3.80, 3.80	□OH -
Val 38	8.01	4.11	2.06	□CH□ 0.86, 0.86
Asp 39	8.27	4.26	2.74, 2.60	□H -
Val 40	7.94	4.09	2.09	□CH□ 0.87, 0.87
Met 41	7.87	4.26	1.94, 1.94	□CH□ -, -; □CH□ 2.04

[a] 1mM in 300mM DPC / 90% H_2O/10% ^2H_2O at 310 K , 20mM MES and pH 5.6

Table S5: $^{1}H^{N}$ and ^{15}N chemical shifts of N-Y4 in water in the presence and absence of DPC micelles.

Res.	δ(^{1}H) [ppm]	δ(^{15}N) [ppm]	δ(^{1}H) [ppm] DPC	δ(^{15}N) [ppm] DPC
M1	-	-	-	-
N2	-	-	-	-
T3	8.171	119.838	8.637	116.741
S4	8.290	114.787	8.413	117.173
H5	8.260	127.348	-	-
L6	-	-	7.652	118.535
L7	8.006	121.818	8.000	116.495
A8	7.953	123.247	7.521	118.535
L9	7.817	119.790	7.395	115.914
L10	7.915	121.832	7.422	115.509
L11	7.897	123.675	7.494	117.061
P12	-	-	-	-
K13	8.257	120.984	8.350	120.579
S14	8.221	117.544	8.319	118.053
P15	-	-	-	-
Q16	8.338	119.671	8.355	119.671
G17	8.243	109.457	8.255	109.367
E18	8.225	120.207	8.232	120.105
N19	8.454	119.731	8.449	119.615
R20	8.252	121.321	8.247	121.173
S21	8.211	116.370	8.214	116.299
K22	8.092	123.461	8.095	123.333
P23	-	-	-	-
L24	8.423	122.478	8.319	121.952
G25	8.315	109.153	8.286	108.684
T26	7.874	115.514	7.910	115.127
P27	-	-	-	-
Y28	-	-	7.886	118.731
N29	8.083	120.162	8.045	119.607
F30	7.989	120.843	8.083	120.695
S31	8.069	116.045	8.084	115.520
E32	8.133	121.685	8.162	121.566
H33	8.288	117.461	8.332	117.816
C34	-	-	-	-
Q35	8.430	121.020	8.460	121.068
D36	8.224	120.984	8.289	121.269
S37	8.096	115.276	8.123	115.324
V38	8.002	120.362	8.008	120.111
D39	8.251	122.983	8.247	122.843
V40	7.918	119.379	7.925	119.118
M41	7.850	128.423	7.831	127.626

Fig. S6: Summary of the meaningful distance restraints as derived from the unambiguously assigned inter-residue NOEs between backbone H^N,H^α and H^β of N-Y4 bound to SDS micelles.

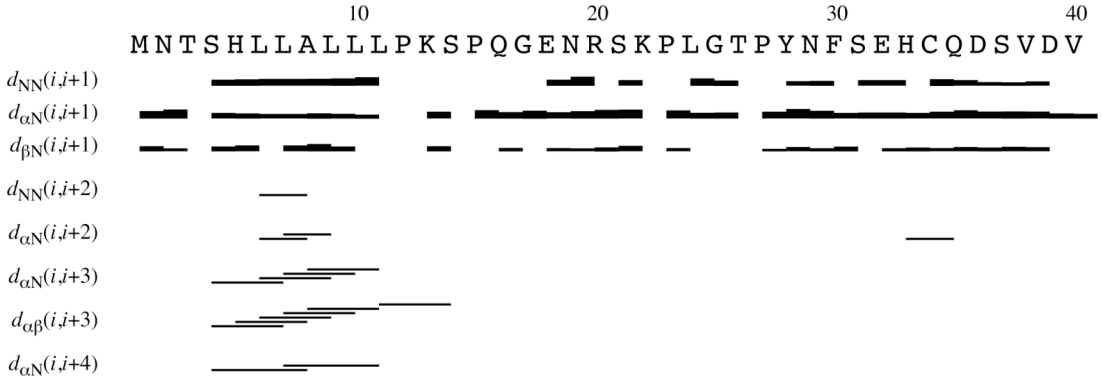


Fig. S7: Comparison of [^{15}N,^{1}H]-HSQC data of N-Y4 in the presence of DPC (left) or SDS (right) micelles:

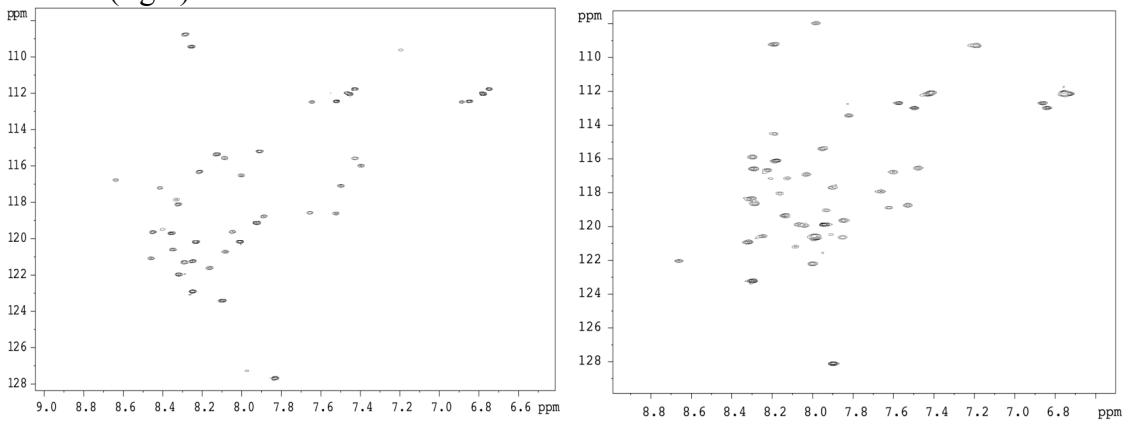


Fig. S8: Comparison of ^{15}N{^{1}H}-NOE data of N-Y4 in the presence of DPC (red diamonds) or SDS (black circles) micelles:

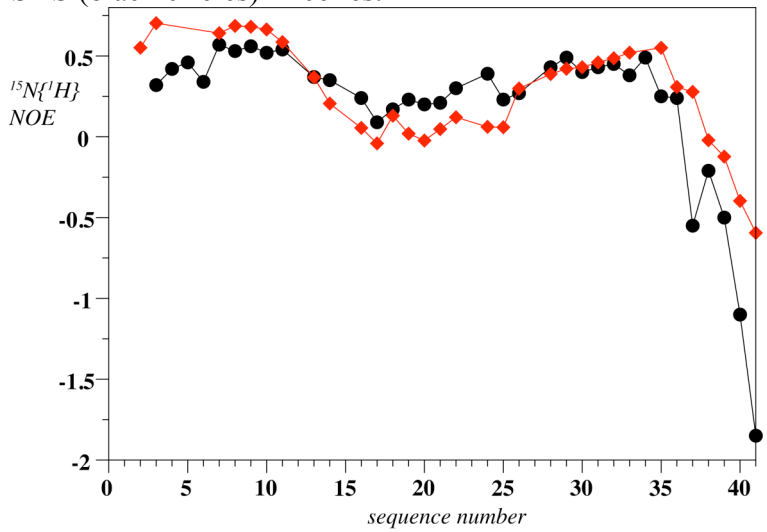


Fig. S9: Left: Residual signal intensity of cross peaks in the [^{15}N,^{1}H]-HSQC in the presence of 5-doxylstearate relative to those in the absence of the spinlabel. Right: Free energies of transfer for whole amino acids from bulk aqueous solution into the water-membrane interface. Values were taken from Wimley et al.[36]

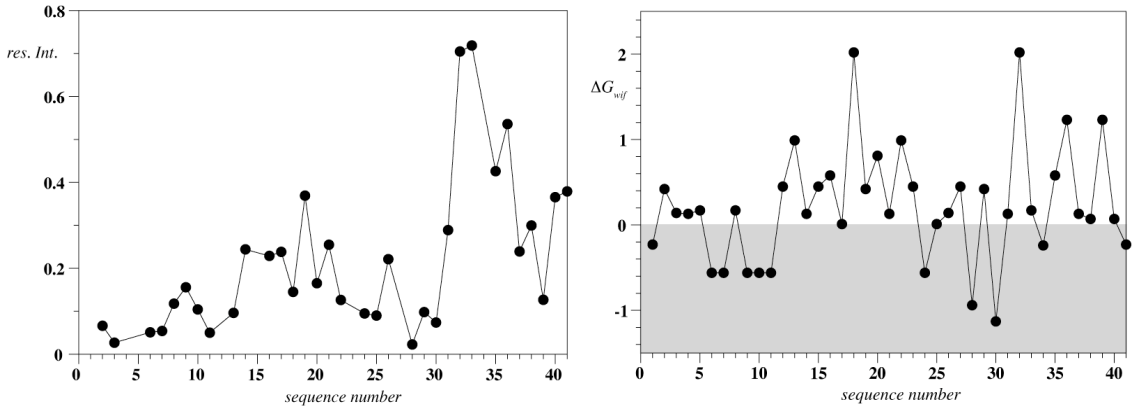


Fig. S10: Left: Overlay of NOESY spectra of N-Y4 with the construct containing the C-terminal lipid attachment (green) (all spectra recorded in the presence of DPC micelles). Right: Expansion showing the assignment of the helical segment of N-Y4.

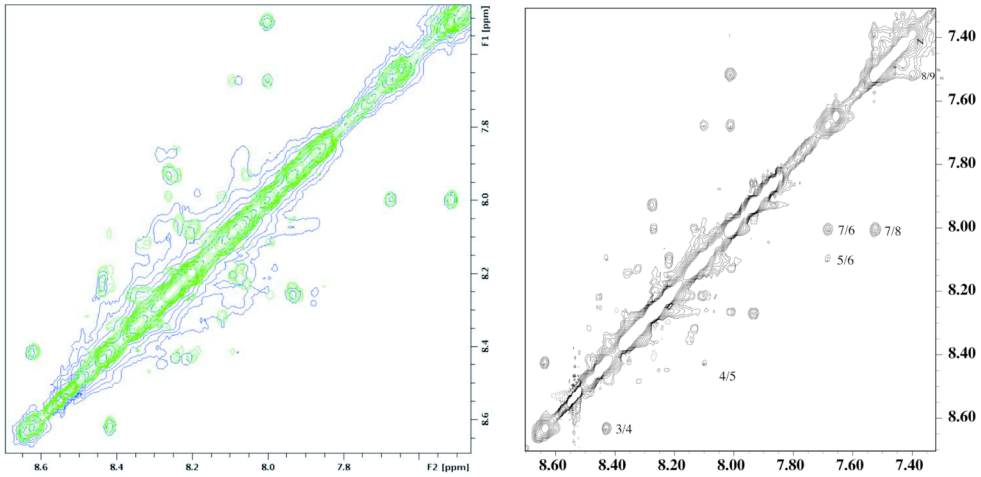


Table S11: Chemical shifts for N-Y4 bound to SDS micelles[a] referenced to the signal of residual HDO at 4.63 ppm

Residue	H^N	H'^{α}	H^{α}	Others
Met 1	-	4.19	2.19	γCH_2 2.55; εCH_3 2.19
Asn 2	8.66	4.98	2.98 , 2.79	δNH_2 7.57, 6.86
Thr 3	8.28	4.26	4.04	γCH_3 1.22; βOH -
Ser 4	8.28	4.23	3.92	βOH -
His 5	8.06	4.57	3.3	δ^1NH -; ε^2H 7.34; ε^1H 8.64; δ^2NH -
Leu 6	7.84	3.99	1.69, 1.78	γH 1.57; δCH_3 0.87, 0.94
Leu 7	8.03	3.95	1.57, 1.78	γH 1.58; δCH_3 0.87, 0.93
Ala 8	7.52	4.05	1.47	
Leu 9	7.47	4.14	1.70, 1.87	γH 1.54; δCH_3 0.86, 0.93
Leu 10	7.59	4.22	1.75, 1.80	γH 1.55; δCH_3 0.82, 0.86
Leu 11	7.65	4.37	1.77	γH 1.56; δCH_3 0.90, 0.94
Pro 12		4.45	2.35, 2.05	γCH_2 2.0, δCH_2 3.54, 3.77
Lys 13	7.90	4.44	1.81	γCH_2 1.46; δCH_2 1.63 ; εCH_2 - ; ζNH_3 6.96
Ser 14	7.95	3.88	3.80	βOH -
Pro 15		4.46	2.31, 1.99	γCH_2 2.06; δCH_2 3.80, 3.75
Gln 16	8.30	4.31	2.15, 1.97	γCH_2 2.30, 2.37; εNH_2 7.42, 6.72
Gly 17	8.18	3.93		
Glu 18	8.14	4.30	2.08, 1.94	γCH_2 2.33; γH -
Asn 19	8.30	4.71	2.84, 2.71	δNH_2 7.49, 6.83
Arg 20	8.0	4.35	1.86, 1.78	γCH_2 1.64; δCH_2 3.20; εNH 7.2; ηNH_2 -
Ser 21	8.22	4.43	3.83	βOH -
Lys 22	7.99	4.35	1.87	γCH_2 1.45; δCH_2 1.75 ; εCH_2-; ζNH_3 -
Pro 23		4.44	2.30, 1.98	γCH_2 1.91; δCH_2 3.77, 3.68
Leu 24	8.03	4.34	1.72, 1.68	γH 1.58; δCH_3 0.94, 0.89
Gly 25	7.98	4.00, 3.91		
Thr 26	7.82	4.61	4.17	γCH_3 1.21; βOH -
Pro 27		4.39	2.17	γCH_2 2.17; δCH_2 3.58
Tyr 28	7.61	4.34	2.71, 2.67	δH 6.88; εH 6.71; ηOH -
Asn 29	7.93	4.68	2.80, 2.63	δNH_2 7.42, 6.76
Phe 30	8.08	4.43	3.20, 3.07	δH 7.28; εH 7.21; ηH 7.07
Ser 31	8.18	4.31	3.93, 3.92	βOH -
Glu 32	7.85	4.23	2.03, 1.91	γCH_2 2.23; γH -
His 33	8.12	4.62	3.34, 3.20	δ^1NH -; ε^2H7.31; ε^1H -; δ^2NH -
Cys 34	8.15	4.66	3.22, 2.98	βSH -
Gln 35	8.24	4.33	2.11, 1.98	γCH_2 2.34; εNH_2 7.41, 6.75
Asp 36	8.31	4.65	2.75, 2.69	βH -
Ser 37	8.18	4.45	3.85	βOH -
Val 38	8.00	4.14	2.09	γCH_3 0.89, 0.85
Asp 39	8.30	4.62	2.60	βH -
Val 40	7.94	4.12	2.09	γCH_3 0.90
Met 41	7.89	4.34	1.92	γCH_2 2.34; εCH_3 -

[a]1mM in 300mM SDS / 90% H_2O/10% ^2H_2O at 310 K, 20mM MES and pH 5.6

Selective N-terminal biotinylation: Proof by MS-MS

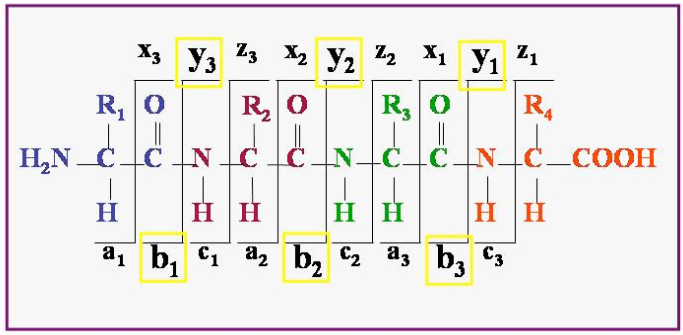
A P L K P E Y P G D N A T P E QMA Q Y A A E L R R Y I N M L T R P R Y-NH_2

MW=4224.8 Da

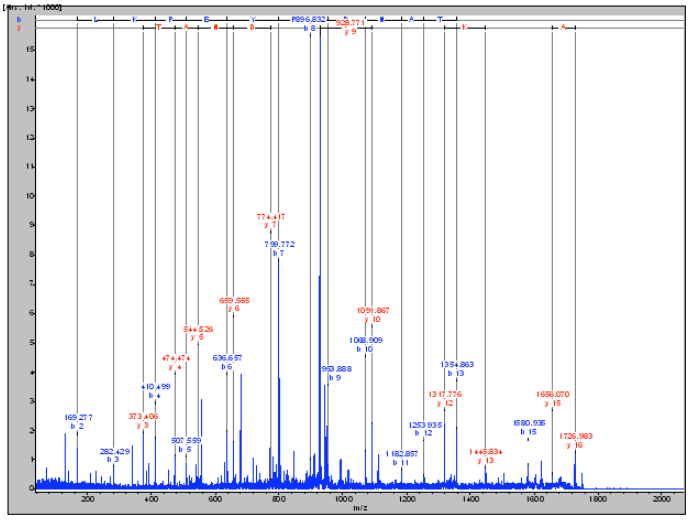
Biotin-(PEO)4-NHS-propionate (NHS-(PEO)4-Biotin): MW=588.67 Da, mass added to the target: 474.6.Da

Pepsin was used to digest the biotinylated peptide yielding a fragment corresponding to biotinylated 1-16 N-Y4. TOF-TOF MS was utilized to analyze this fragment:

The scheme of peptide fragmentation

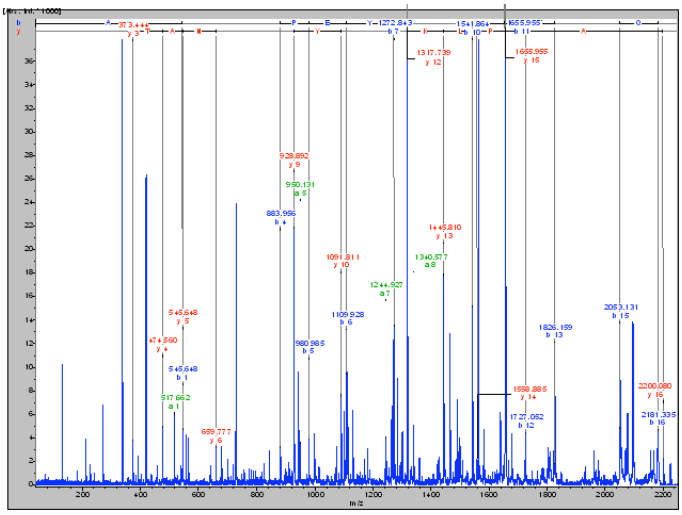


The non-biotinylated sample served as a control, and the TOF-TOF result is as follows:



	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16
Expected	372.3	473.4	544.5	658.6	773.7	830.8	927.9	1091.1	1220	1317.3	1445.5	1558.6	1655.7	1726.8
Observed	373.4	474.5	544.5	659.5	774.4		928.8	1091.9		1317.8	1445.8		1656.1	1726.9

Almost all expected ions from the digest are observed. Thereafter, the biotinylated sample was fragmented under the same conditions, and the TOF-TOF MS spectrum is depicted below:



If we assume that biotin was coupled to the N terminus, the following expected y ions as listed

	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16
Expected[1]	473.4	544.5	658.6	773.7	830.8	927.9	1091.1	1220	1317.3	1445.5	1558.6	1655.7	2201.1
Expected[2]	473.4	544.5	658.6	773.7	830.8	927.9	1091.1	1220	1317.3	1920.1	2033.2	2130.3	2201.1
Observed	474.6	545.6	659.8			928.9	1091.8		1317.7	1445.8	1558.9	1656.1	2200.1

Expected[1]: expected y ions if biotin was coupled to the N terminus
Expected[2]: expected y ions if biotin was coupled to lysine

The observed and expected y ions have been summarized in the table above. It is evident that the observed y ions, in particular y ions 13-16 correspond to those where biotin was coupled to the N terminus. In addition b1 ions 545.6 can be identified corresponding to biotinylated N terminus residue (MW of biotinilyated Alanine b ion=546.0). None of the fragments occurring only for Lys-biotinylated peptide was observed.

Chapter 3

Properties of the N-terminal domains from Y receptors probed by NMR spectroscopy

3.1 Introduction

G-protein coupled receptors (GPCRs) present the pharmacologically most important class of receptors and the most important target for pharmaceutical drugs[1]. Although significant progress has been made in structural studies of GPCRs, a much more detailed picture still is highly desirable. Nevertheless, the structures of bovine rhodopsin[2], the very recently published data on the b-adrenergic receptor[3; 4] and on squid rhodopsin[5] have improved our understanding of this biologically important class of proteins.

Generally the structure of GPCRs can be described as an extracellular N terminal domain (ranging in size from ten to several thousand residues), which is anchored in the plasmamembrane by 7 transmembrane helices (7TM segment). The latter are interconnected between themselves by three intra- and three extracellular loops. The 7TM segment is followed by a cytoplasmic C terminal domain.

While the extracellular N-terminal domain of bovine rhodopsin was surprisingly well-structured and revealed the non-anticipated presence of a short anti-parallel beta-sheet, the corresponding segment of the b-adrenergic receptor could not be traced in the electron maps presumably because of its inherent flexibility[3; 4]. Previously, we have in detail investigated structural properties of a 41 amino acid fragment corresponding to the N-terminal domain of the Y4 receptor (N-Y4)[6]. This receptor belongs to a class of GPCRs targeted by neurohormones of the neuropeptide Y family[7; 8]. The Y receptors are comprised of four subtypes called Y1, Y2, Y4, and Y5 with Y4 showing high affinity and specificity for the pancreatic polypeptide (PP). While unstructured in solution a short a-helical stretch comprising residues 5 to 11 was observed in the presence of phospholipids micelles for N-Y4.[6] In this work we now report on our recent studies on structural properties of all other N-terminal domains from the human Y receptors. Synthetic routes for recombinant production of the polypeptides in isotopically-labeled form are described and compared to each

other. In general the N-terminal domains from the Y1 and Y4 receptors behave similarly, and the same is true for those from the Y2 and Y5 receptors. Expression of N-Y1 and N-Y4 required fusion to the insoluble protein ketosteroidisomerase, from which it was liberated by enzymatic cleavage using the TEV protease in the presence of a mild detergent. In contrast, N-Y2 and N-Y5 could be expressed as soluble ubiquitin fusions, and cleavage was easily achieved with the help of yeast ubiquitin hydrolase.

The N-terminal domains from all Y receptors are fully unstructured in aqueous solution, as shown by measurements of the internal backbone dynamics. In contrast, in the presence of phospholipid micelles N-Y4 adopts a short a-helix in a segment mainly comprised of hydrophobic residues. N-Y1 is largely helical although remaining flexibility precludes a detailed structural analysis. N-Y5 is segregated into more structured and rather flexible regions, similarly to N-Y4. However, measurements of internal backbone dynamics revealed secondary structure to be less stable than in N-Y4. N-Y2 does not interact with the micelles and remains unstructured also in that environment.

In our previous work we demonstrated that N-Y4 interacts with PP[6]. Surface plasmon resonance (SPR) measurements indicated weak (Kd 50mM) binding, and subsequent mutagenesis experiments revealed that electrostatic interactions from anionic ligand and cationic N-Y4 residues contributed to that interaction. In this work we also tested binding of the principal members of the NPY family (the neuropeptide Y (NPY), the pancreatic polypeptide (PP) and the peptide YY (PYY)) to all other N-terminal domains from this class of GPCRs.

3.2 Results

3.2.1 Expression of N-terminal Domains in Isotopically-labeled Form

The structure of peptides can mostly be solved by relying solely on homonuclear ^{1}H-^{1}H correlation experiments. Such peptides are therefore usually produced by solid phase peptide synthesis (SPPS)[9; 10]. Isotope labeling, however, is required for the study of backbone dynamics using ^{15}N relaxation, and such labeling also facilitates chemical shift mapping experiments for the study of macromolecular interactions. The high cost of ^{13}C- and/or ^{15}N-enriched amino acids usually prohibits the usage of SPPS and necessitates recombinant production. *E.coli* is still the expression system of choice for most proteins because of the ease of its genetic manipulation and because of the ability of *E.coli* to synthesize amino acids from glucose and inorganic ammonia salts serving as the sole sources of carbon and nitrogen, respectively[11].

Since peptides are rapidly degraded in *E.coli*, they are usually expressed linked to a (more) stable fusion partner. The chosen fusion partner should allow the expression of the fusion constructs in high amounts and it should allow the specific separation of the fusion partner and the peptide after the purification of the fusion construct[12]. The first aim can usually be achieved by selecting a protein as fusion partner, which itself can be produced in high yields in *E.coli*. Specific cleavage from the fusion partner can be accomplished for systems for which a specifically hydrolase is available (method 1), or by introducing a unique cleavage site between the fusion partner and the peptide sequence of interest (method 2) (see Fig. 1). Such a scission site can be either an amino acid sequence specifically recognized by a protease (method 2a) or a site prone to chemical cleavage (method 2b).

The most convenient strategy in terms of workload is usually method 2b. In this method high yield is achieved by fusing the peptide to a highly water insoluble protein which will lead to the accumulation of the fusion construct in inclusion bodies[11]. Inclusion bodies already contain the target protein at high concentrations, and typically require only very few additional steps of purification. In case of cyanobromide cleavage very high efficiencies have been reported[13], but the target sequence is not allowed to contain Met residues. Other methods such as hydroxylamine cleavage[14], are much less efficient, and often result in further chemical modifications of the target peptide. Methods that use enzymatic cleavage (methods 1

and 2a) require that the fusion protein is soluble under conditions that are compatible with enzymatic activity.

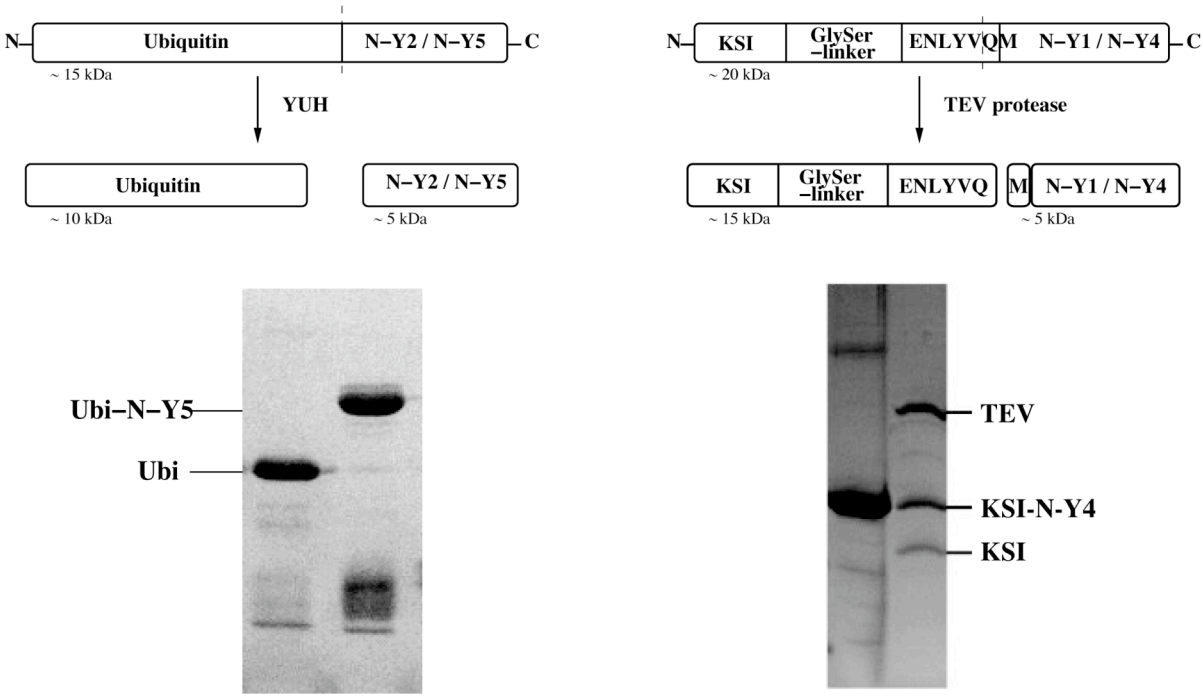


Figure 1: Scheme showing the two strategies used to produce peptides corresponding to the N-terminal domains of the Y receptors in isotope-labeled form.

Since the four Y receptor N-terminal fragments studied herein are all reasonably water-soluble and contain Met residues we initially decided to express them in ^{15}N-labelled form as C-terminal fusions to N-terminally decahistidine-tagged yeast ubiquitin[15]. After purification of the fusion construct by Ni-affinity chromatography the desired peptide was liberated with a hexahistidine-tagged yeast ubiquitin hydrolase (YUH). This system allowed the recovery of about 6 mg of ^{15}N-labelled N-Y2 and N-Y5 from 1 L of culture. Unfortunately, attempts to express N-Y1 and N-Y4 using this method resulted in unspecific C-terminal degradation (see supplementary Table S10). To circumvent intracellular proteolysis, N-Y1 and N-Y4 were expressed as a fusion to the highly water-insoluble protein ketosteroidisomerase (KSI), which resulted in accumulation of the fusion protein in inclusion bodies. A TEV protease cleavage site was introduced between KSI and the target peptide[16; 17]. The sequence recognized by the TEV protease is ENLYFQ with Q as the P1' residue. To achieve the natural peptide sequence after cleavage, the P1' residue was replaced with the first residue from the target sequence (here it is Met)[16], and an additional GSGSGS linker was inserted to prevent steric hindrance during cleavage.

A problem of the chosen strategy was that the water-insoluble fusion protein must be solubilized in detergent that is compatible with activity of the TEV protease[18]. After extensive detergent screening, we observed that the ionic detergent sarcosyl solubilizes the fusion protein while preserving TEV protease activity to a satisfactory extent (see Table S10). Cleavage efficiency for this system is around 40% allowing recovery of about 2 mg of ^{15}N-labelled N-Y1 and N-Y4 from 1 L of bacterial culture.

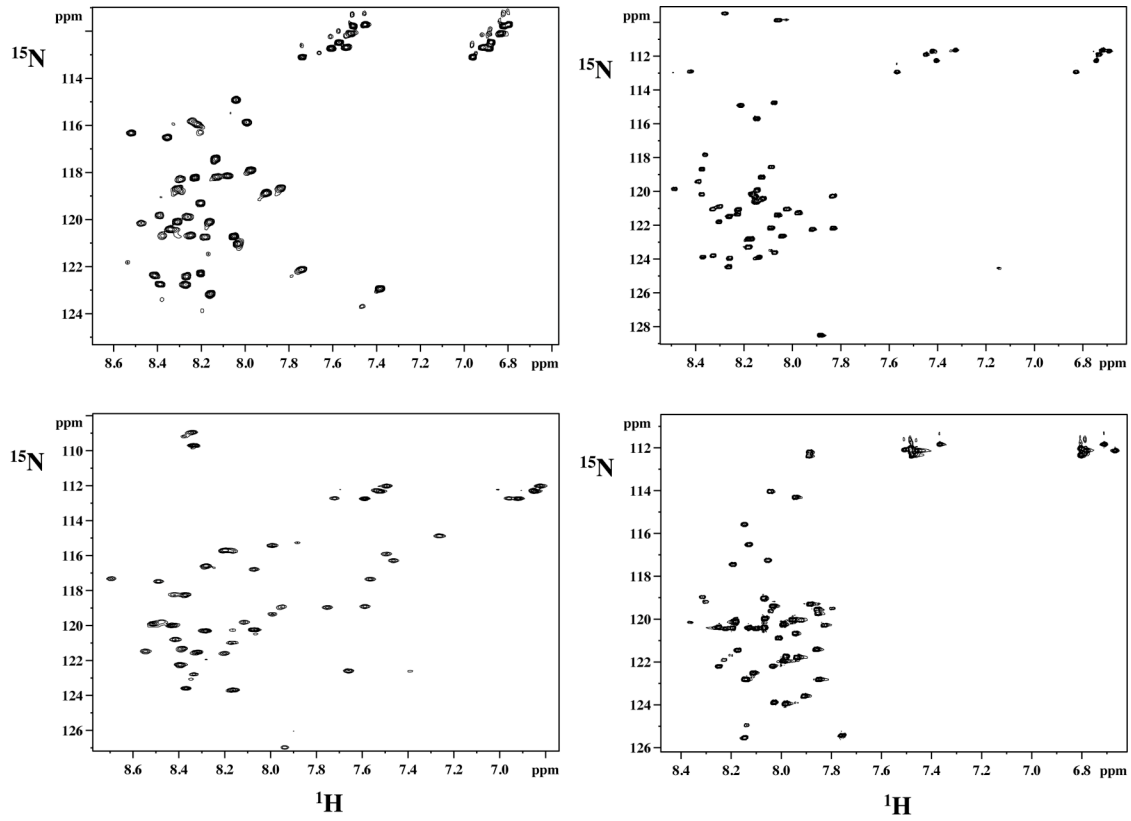


Figure 2: [^{15}N,^{1}H]-HSQC spectra of all Y receptor N-terminal domains, recorded at 310 K in the presence of DPC micelles. Top left: N-Y1, top right: N-Y2, bottom left: N-Y4, bottom right: N-Y5.

3.2.2 Assignment of Chemical Shifts

Sequence-specific resonance assignments were done using the strategy developed by Wüthrich and coworkers[19]. In this strategy spin systems are assigned by experiments based on scalar couplings (e.g. COSY-type or TOCSY-type transfer) and NOEs are used to link them in sequential order. Due to extensive resonance overlap of the poorly folded peptides ^{15}N-resolved three-dimensional TOCSY or NOESY data had to be utilized for this task. Representative [^{15}N,^{1}H]-HSQC spectra of all four peptides are depicted in Fig. 2. In case of N-Y5 a ^{13}C,^{15}N-labelled sample, allowing the acquisition of triple resonance spectra, was required. For N-Y1, a set of

experiments was first recorded in aqueous buffer. After completed analysis in water the assignments were transferred and adjusted to the spectra recorded in the presence of DPC micelles with the help of NOESY spectra. A set of tables containing complete assignments for all the proton-nitrogen correlation maps, as well as almost all proton chemical shifts of N-Y1 can be found in the supplementary materials S1-S5

3.2.3 Screening Structural pProperties Ssing ^{15}N Relaxation and CD Spectroscopy

CD spectroscopy is a convenient tool to estimate the type and content of secondary structure in peptides and proteins. The CD spectra of all N-terminal domains in the presence of DPC micelles are depicted in Fig 3. The spectrum of N-Y2 displays its maximum around 197 nm, the typical absorption band of unstructured peptides. For all other peptides the main band is red-shifted and indicates population of helical substructures. The intensities of the absorptions, however, also clearly show that the helical content is very low in all cases, and the typical bands at 208 and 222 nm are not visible. For N-Y4, for which we previously observed an a-helix involving residues 4 to 11, the absorption is stronger than for the other peptides.

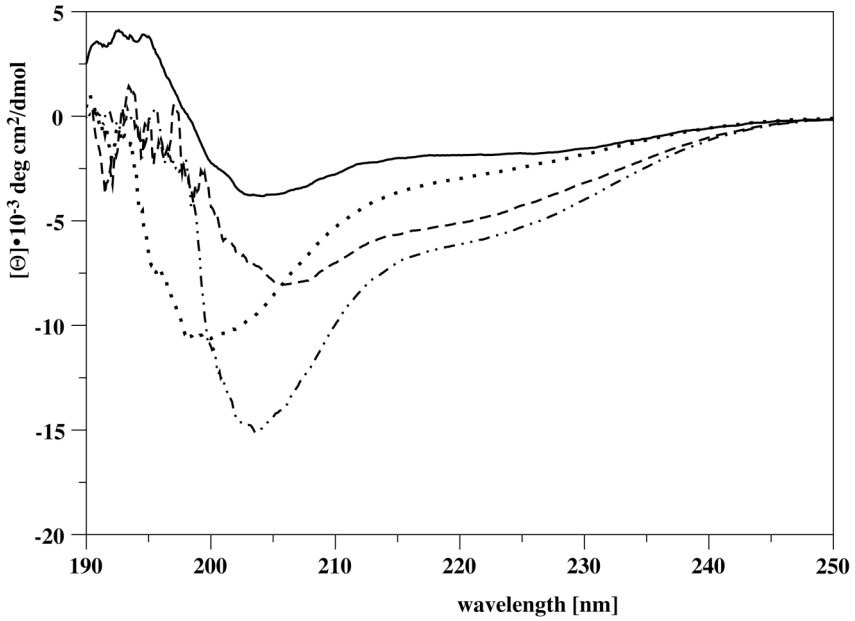


Figure 3: CD spectra of peptides from all N-terminal domains, recorded at 37 °C in 300 mM DPC, 20 mM MES pH 5.6 solution. Data are shown for N-Y1 (solid line), N-Y2 (dotted line), N-Y4 (dash-dotted line) and N-Y5 (dashed line). Data are converted to molar ellipticities.

The dispersion of the NMR signals in the region of the amide protons is traditionally used to estimate to which extent a peptide or protein is folded[20]. In case

of the N-terminal domains from the Y receptors signal dispersion of all peptides was small, indicating that they were largely unfolded. To better access whether these peptides still contained folded segments we recorded the ^{1}H-$^{15}N\{^{1}H\}$-NOEs (H-NOEs). These values range from 0.6 and 0.8 for well-folded elements of secondary structures, and progressively decreases for more flexible amide moieties resulting in negative values for fully flexible segments[21]. The H-NOE data for all N-terminal peptides reveal that all peptides are essentially unstructured in aqueous buffer (data not shown).

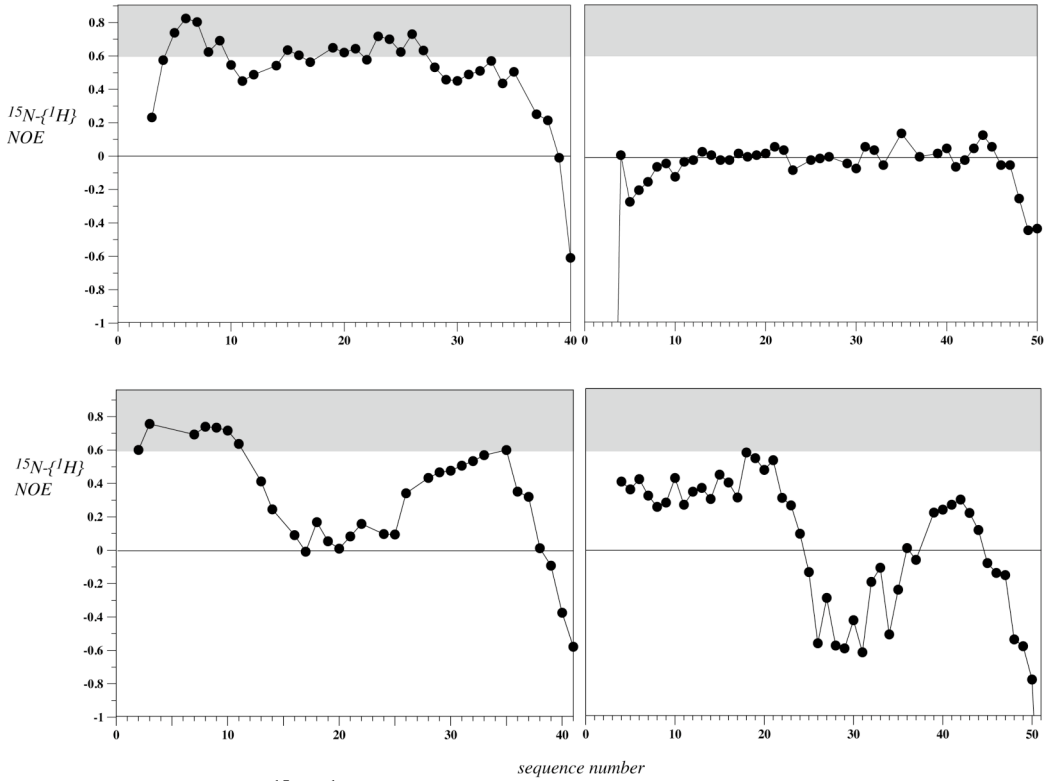


Figure 4: Values of the ^{15}N-$\{^{1}H\}$NOE, recorded at 700 MHz proton frequency along the sequence for Y1 (top left), Y2 (top right), Y4 (bottom left) and Y5 (bottom right). The area containing values larger than 0.6, indicating rather wellfolded segments, has been shaded in gray.

Since in the naturally occurring GPCR the N termini are attached to a membrane-protein the backbone dynamics were additionally probed in the presence of a commonly used membrane-mimicking detergent, dodecylphosphocholine (DPC)[22] (see Figure 4). Again the peptides are not rigidly structured. In the case of N-Y4 we could previously show that a rather stable hydrophobic a-helix is formed between residues 4 and 11, present both in zwitterionic (DPC) as well as in anionic (SDS) micelles[6], reflected by H-NOEs exceeding values of 0.6. In contrast, the N termini from all other Y receptors are less well ordered.The N-Y2 is fully flexible most likely

due to the complete lack of interactions with □hospholipids surfaces. The absence of such contacts is supported by the fact that essentially no chemical shift changes occur between N-Y2 in aqueous buffer and in DPC micelles. In contrast, both N-Y1 and N-Y5 reveal short stretches of the polypeptide chain that become rigidified in the presence of the micelles.

3.2.4 Structures of the N-terminal Domains in Presence of Phospholipid Micelles

The NOE data of N-Y4 revealed the presence of a hydrophobic helix in the segment comprising residues 5 to 11. In addition, a nascent helix was observed in the region including residues 26 to 35. Inspection of the H-NOE data depicted in Figure 4 clearly indicates that N-Y2 is devoid on any structured segments. Moreover, the H-NOE of N-Y5 is generally below 0.6 and mostly values are even smaller than 0.4. In our experience secondary structure cannot reliably be determined in these cases. We speculate that the molecule, similarly to N-Y4, is segregated into a N-terminal helical region, and a much more destabilized shorter C-terminal helical region separated by a longer non-ordered segment, but the peptide is not ordered sufficiently well to allow for structural characterization by NMR in detail.

In case of N-Y1, however, elevated values of the H-NOE are observed indicating that this polypeptide may be amenable to more detailed structural studies. Accordingly, we have assigned all proton and nitrogen resonances of N-Y1. Little chemical shift dispersion of amide proton resonances complicated the assignment process, and use of 3D ^{15}N-resolved NOESY or TOCSY spectra had to be made. During assignment a larger number of contacts involving sequential amide protons were observed, indicating that the fy space of helical backbone conformations was significantly populated. Such stretches were for example observed involving residues 4 to 9 and residues 24 to 32. An expansion of the spectral region of the [^{1}H,^{1}H]-NOESY that displays the sequential amide proton NOEs in the segment from 24 to 32 is shown in Figure S6 in the supplementary materials. However, except for two a,N (i,i+3) NOEs observed in the segment 4-9 no medium-range contacts were found. The relative strength of intra-residual and sequential aH,NH contacts changes between extended and helical conformations[23], with the intra-residual distance in helices stronger than the sequential one, whereas in extended or unfolded segments the sequential distance is much shorter. A comparison of peak intensities revealed that the

sequential NOEs were generally stronger, and in the light of sequential contacts of amide protons, indicate conformational averaging between helical and extended conformations to some extent. Considering this observation it was not really surprising that persistent violations remained in the structure calculations, and helical conformations were only seen involving residues 4 to 9, a region, in which the H-NOE is larger than 0.6. The ^{3}J(HN,Hạ couplings were larger than 6.5 Hz throughout the sequence (data not shown), reflecting the remaining conformational instability of N-Y1. To our surprise we have not been able to detect any medium range contacts in the segment 15 to 28, which according to the dynamics data should also be better ordered. We suspect this region to be transiently helical considering the occurrence of sequential amide proton contacts throughout this segment.

To summarize, the spectroscopic data indicate that N-Y4 and N-Y5 are similar in that both contain two helical regions separated by a flexible central segment, with only the N-terminal helix in N-Y4 being well ordered. N-Y1 is largely helical between residues 4 and 28, but the remaining conformational flexibility precludes its detailed structural analysis. N-Y2 is fully flexible and devoid of any detectable residual structure.

3.2.5 Interaction Studies with Neuropeptides from the NPY Family

We have recently proposed that the peptides of the NPY family may transiently bind to the N-terminal domains in order to become transferred from the membrane-bound state into the genuine binding pocket of the receptor[6; 24]. While in that work surface plasmon resonance was used to establish the strength of the bPP-NY4 interaction, preliminary experiments using bPP or pPYY and the N-terminal domains from the other receptors have indicated that the interaction between the peptides and the other N-terminal domains are too weak to be detected by SPR. In this work we have now utilized chemical shift mapping experiments both in the presence and absence of DPC micelles in order to derive preliminary data on binding of the peptides from the NPY family to N-Y1, N-Y2 and N-Y5. The changes of chemical shifts of the neuropeptides upon adding 2 equivalents of the N-terminal domains are summarized in Figure 5. First of all we noticed that the magnitude of the changes is larger in the presence of DPC micelles as compared to aqueous buffer. This was surprising considering that in case of N-Y4 the interaction in the absence of micelles

was stronger than in the presence as judged by SPR[6]. It should be mentioned, however, that changes in chemical shift mapping depend on both the population difference between bound and free species, but also on the type of structural changes. Contacts with aromatic residues for example will for instance result in larger changes than contacts with polar side chains. Detachment of peptides from the micelle surface, possibly required for binding to the receptor N termini, will transfer part of the residues into a completely different environment, and we believe this effect to account for the larger changes observed in DPC micelles.

In the presence of lipids, chemical shift changes for bPP are much larger than for the other peptides. When adding N-Y1 or N-Y5 to bPP the same residues of bPP are affected in a similar manner. In all cases the largest changes occur in the N-terminal half of bPP or in the segment from residues 26 to 35. In case of pNPY changes are only appreciable for residues 14 and 15 upon addition of N-Y5. Changes in pPYY are very similar for all the N-terminal domains, and largest for residues 21, 26 and 27.

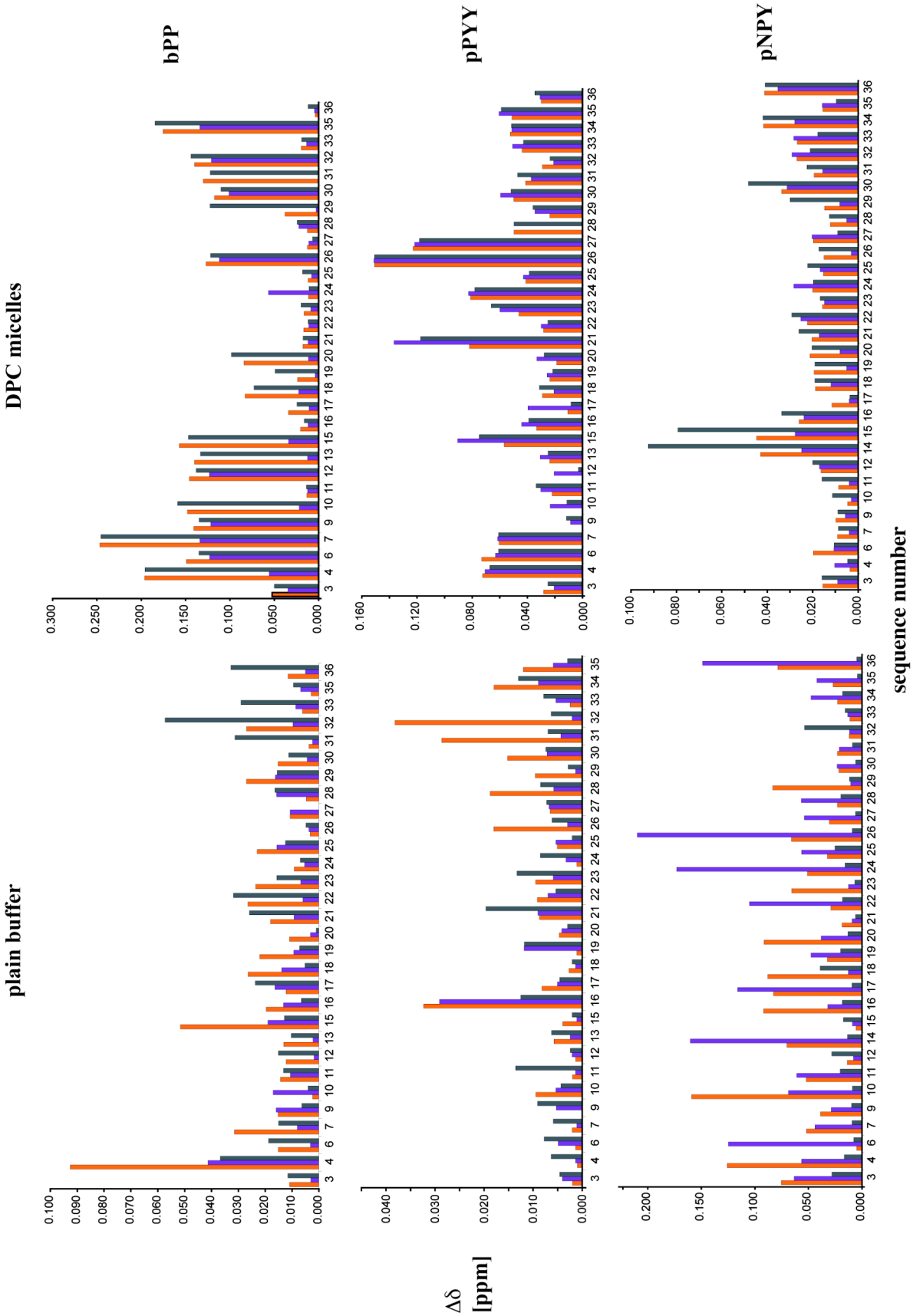


Figure 5: Chemical shift deviation of bPP (top panel), pPYY (middle panel) and pNPY (bottom panel) upon addition of N-Y1(orange bars), N-Y2 (pink bars) and N-Y5 (gray bars) in aqueous buffer (left column) or in the presence of DPC micelles (right column).

In the absence of lipids, larger changes are detected for pNPY throughout the sequence upon adding N-Y1 or N-Y2, but are fairly small for bPP and pPYY, except for residue 4 of bPP upon addition of N-Y1. Pronounced differences are observed for pNPY residues 6, 14, 24, 26 and 36, and for residues 4 and 10 when adding N-Y1 or N-Y2, respectively. We like to emphasize here that the spectra of bPP and pPYY in water are concentration independent whereas the oligomerization state of pNPY is strongly concentration dependent[25; 26; 27]. In addition, the spectrum of pNPY is highly dependent on pH and other environmental variables. Although we tried to control these as tightly as possible we cannot fully exclude that part of the observed changes may relate to issues not directly linked to binding.

To investigate whether pNPY really associates with N-Y2 we have performed a titration experiment, in which up to 10 equivalents of pNPY were added to ^{15}N-labelled N-Y2 (see Fig. 6). The data clearly show concentration-dependent changes of positions of resonances from the N-Y2. Resonances in the segments comprising N-Y2 residues 16-21 and 33-50 are mostly affected. The data point to a low-affinity interaction of pNPY towards N-Y2 without much specificity.

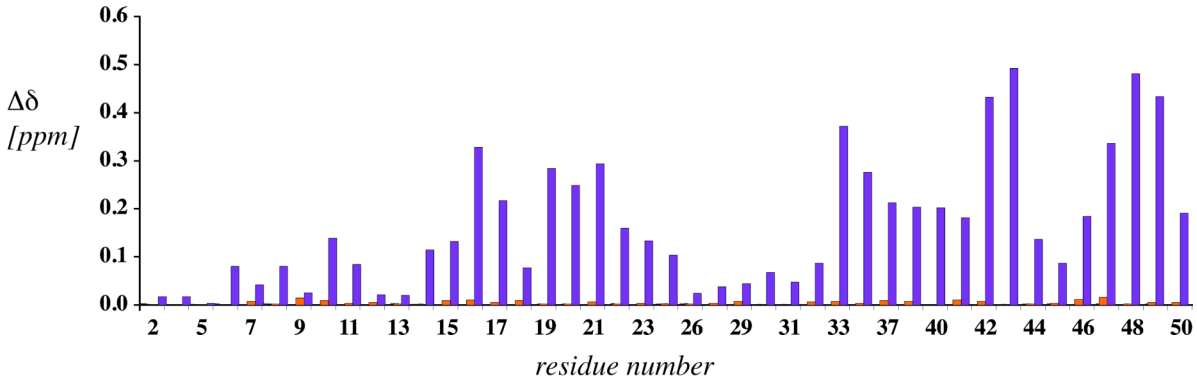


Figure 6: Chemical shift deviation of N-Y2 after addition of 1 and 10 equivalents of pNPY (from left to right). For additional data points at 0.5, 2 and 4 equivalents see Figure S12.

To summarize the interaction studies we can say that significant and reliable effects were only detected in the presence of DPC micelles and that all peptides interact with the three N-terminal domains. The interaction with bPP results in the largest changes.

3.3 Discussions

We have postulated that binding of ligands to Y receptors is preceded by association of the ligands to the plasma membrane. Thereby, the apparent concentration of the ligand is increased and the search for the receptor reduced from three to two dimensions [28; 29]. We now studied whether parts of the receptor that protrude into the extracellular compartment may help in transferring ligands, which have accumulated in vicinity of the membrane, into the binding pocket. Such portions of receptor that point into the extracellular space are the N-terminal domains. Herein, we have developed strategies to produce these polypeptides recombinantly in isotopically-enriched form for use in high-resolution NMR studies.

The work has demonstrated that these peptides can all be expressed as soluble fusions to ubiquitin. However, N-Y4 and N-Y1 are degraded in the intracellular milieu, and hence much better yields were obtained using insoluble fusions. Cleavage of the target sequence from the insoluble fusion partner could be obtained by solubilizing the fusion protein in the mild detergent sarcosyl, which proved to be compatible with enzymatic activity of the TEV protease used to cleave the peptide from the fusion protein.

Studies on the structure and dynamics of the peptides using NMR revealed that they are all completely disordered in aqueous buffer. In the presence of hospholipids micelles, segments of most receptor N termini became conformationally stabilized, with the exception of N-Y2, which remained. Otherwise, more (N-Y4) or less stable (N-Y1 or N-Y5) helical segments occurred within the sequences. For all N-terminal peptides chemical shit changes occurred in the presence of DPC micelles, except for N-Y2. This implies that all other peptides associate with the micelle to some extent. Previously, we have made extensive use of the thermodynamic data of Wimley and White for partitioning of single amino acids into the water-membrane interface or the membrane interior[30] to rationalize how peptides interact with phospholipids micelles. A common observation was that the occurrence of the aromatic residues Trp and Tyr help in anchoring peptides in the interface[31]. The partitioning values of the four sequences of the N-terminal domains from the Y receptor subtypes are shown in Figure S7 in the supplementary materials. In N-Y4 a stretch comprising residues 5 to 11 is predicted to show partitioning into the micelle interior. This corresponds exactly to the region that becomes helically structured in

the presence of micelles. In case of N-Y2 many negatively charges residues occur throughout the sequence, while they are clustered in the central (unstructured) segment in N-Y4. Even more importantly, many Pro residues are present in N-Y2 that might prevent formation of secondary structure. The sequence of N-Y5 in comparison to N-Y2 is much more amphiphilic in nature, and therefore more likely to favorably interact with the micelles. Again, the regions that become better structured in the presence of DPC micelles correspond to stretches rich in hydrophobic/aromatic residues and hence are predicted to partition into the micelles.

Our interaction studies using chemical shift mapping indicated that bPP strongly interacts with all N-terminal domains, but differences in the sensitivity of certain positions are observed. In contrast, for pPYY the changes are smaller, but more uniform upon addition of the different N termini. For pNPY the changes are very small, although the sequence of pNPY displays more than 80% homology to the one of pPYY. The difference between pNPY and pPYY is very intriguing. We speculate that this difference may be due to structural transitions occurring in pPYY when transiently associating with the N-terminal domain. pPYY in aqueous buffer displays the so-called PP-fold[32], and tertiary structure is lost when pPYY associates with the membrane surface, or is transferred into a more hydrophobic solvent (e.g. methanol)[33]. Transient binding of PYY to any of the N-terminal domains is expected to alter the equilibrium between membrane-associated peptide, which is devoid of tertiary structure, and the membrane-detached peptide, that could possibly re-adopt its PP fold. Such a change will result in large chemical shift changes and would explain why the differences in pPYY are so much larger than for pNPY upon addition of the different N termini. In addition the changes due to such a structural transition may be much larger than those due to direct contacts, and therefore can possibly explain, why the changes in pPYY are so uniform across the different N-termini. Considering the sensitivity of the chemical shift mapping methodology to structural changes, but on the other hand the inherently low sensitivity of NMR, we consider results from SPR measurements to be more reliable for quantifying (but not for detecting) such interactions. In our BiaCore measurements we could detect strongest binding (Kd approx. 50 mM) for the bPP-NY4 interaction, and chemical shift mapping also revealed the largest changes for bPP upon addition of N-Y4.

To summarize this work has described synthetic methods to produce all N-terminal domains in isotopically labeled form in quantities sufficient, for the analysis

by various biophysical methods. Structural studies revealed them to be fairly flexible. However, while N-Y2 is fully unfolded residual helical structures were detected in N-Y1 and N-Y5. For the case of N-Y4 we could previously detect a short rather rigid a-helical stretch in the presence of DPC micelles. In contrast to N-Y4, the nascent helical regions of N-Y1 and N-Y5 contain too much residual motion, so that structure calculations did not fully converge towards a-helical structures. All peptides interact with the N terminal domains of N-Y1, N-Y2 and N-Y5, but the interactions are weaker than those previously described for bPP binding to N-Y4.

3.4 Materials and Methods

3.4.1 Materials

$^{15}NH_4Cl$ was from Spectra Isotopes (Columbia, USA), d_{38}-DPC- (99%-d), and D_2O was from Cambridge Isotope Laboratories (Andover, Massachusets, USA). 5-doxylstearic acid was from Aldrich (Buchs, Switzerland). Oligonucleotide primers were synthesized by Microsynth GmbH (Balgach, Switzerland).

3.4.2 Expression and Purification of N-terminal Domains

Depending on their stability against proteolysis the N-terminal domains were either expressed as fusions to ubiquitin (N-Y2 and N-Y5) or to ketosteroidisomerase (N-Y1 and N-Y4).

In case of N-Y2 and N-Y5 the amino acid sequence was reverse translated into a DNA sequence taking into account the preferred *E.coli* codon usage including a terminal stop codon and a *Sal*I restriction site. The resulting fragments were purified by electrophoresis and gel extraction and digested with *Sal*I, resulting in fragments that were then blunt-ended on one side and contained *Sal*I-cohesive end on the other end. These fragments were ligated into the pUBK19 vector (gift from T. Kohno, Mitsubishi Kasei Institute of Life Science, Tokyo, Japan), which had been digested with *Nsi*I and *Sal*I and purified before. The resulting plasmids were sequenced and transformed into C41 cells[34]. For production of ^{15}N-labeled peptides M9 minimal media containing ^{15}N-ammoniumchloride as the sole nitrogen source was used, otherwise expression was done on LB medium. In each case 1 liter of medium containing 50 mg/ml kanamycin was inoculated with 10 ml of an overnight LB culture. Cultures were induced at OD_{600} around 0.5 with 0.4 mM IPTG. LB- and minimal medium cultures were grown under induction for 4 h and 11 h, respectively. Cells were harvested by centrifugation on a Sorval GSA rotor at 4 °C and stored at -20 °C. The cell pellets were thawed on ice for 1 h and resuspended in 25 ml denaturing basic buffer (50 mM Tris pH 8, 6 M GdnHCl, 100 mM NaCl, 1 mM b̃mercaptoethanol). The suspension was lysed by sonication on ice.

The ubiquitin fusion proteins were purified by Ni-NTA chromatography. Refolding was achieved by applying a linear gradient to exchange the denaturing basic buffer to native binding buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM b-mercaptoethanol, 20 mM imidazole), and the protein was eluted with binding buffer

containing 200 mM imidazole. The eluates were diluted 10-fold with basic buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM b-mercaptoethanol) and a 1 mg/ml YUH-solution (for expression and purification of YUH see supplementary S8) was added in a 20-fold dilution. The cleavage reactions were allowed to proceed for 3 hours at 37 °C.

In case of N-Y1 and N-Y4 the DNA sequences were subcloned from wt cDNA of the corresponding Y receptor (University of Missouri-Rolla (UMR) cDNA Resource Center by PCR). During PCR, a GSGSGS linker followed by TEV cleavage sequence was introduced at the N terminus of the target sequence. After digestion with XhoI and EspI, the fragments were ligated with T4 DNA ligase into the pET31b vector, which had been digested with XhoI and EspI. The correctness of the constructs was verified by DNA sequencing (Synergene Biotech, Switzerland). The resulting plasmids were transformed into BL21(DE3) for expression. For production of ^{15}N-labeled peptides M9 minimal media containing ^{15}N-ammoniumchloride as the sole nitrogen source was used, otherwise expression was done in LB medium. In each case 1 liter of medium containing 50 mg/ml kanamycin was inoculated with 10 ml of an overnight LB culture. Cultures were induced at OD_{600} of 0.7 with 1 mM IPTG, harvested after 5 hours by centrifugation on a Sorval GSA rotor at 4 °C and the pellets were stored at -20 °C.

The fusion proteins were purified from inclusion bodies by Ni-NTA chromatography in presence of 6 M GdnHCl. After removal of GdnHCl by dialysis the precipitated fusion protein was solubilized in 50 mM Tris pH 8.0 in the presence of 2% N-lauryl sarcosine upon sonication to a final concentration of 2 mg/ml. The resulting solution was dialyzed against a 20-fold excess of 50 mM Tris, pH 8.0 for 4-6 times. The solution was diluted 10 times with 50 mM Tris pH 8.0 and EDTA and DTT were added to a final concentration of 0.5 mM and 1 mM, respectively. TEV protease (for expression and purification of TEV protease see supplementary materials S9) was added to a final concentration of 100 mM and the cleavage mixture was incubated at 4°C over night.

All target peptides were finally purified by C18-RP-HPLC (Vydac, USA) by using a water/acetonitrile/0.1% TFA gradient. Yields ranged from 3 mg to 20 mg peptide from 1 liter of culture. The mass of all peptides was confirmed by MALDI-TOF MS or ESI MS: N-Y1: 4532.9 Da (theoretical value: 4533.0 Da); ^{15}N-N-Y1: 4587.0 Da (theoretical value: 4587.0 Da); N-Y2: 5509.3 Da (theoretical value: 5510.0

Da); ^{15}N-N-Y2: 5568.0 Da (theoretical value: 5570.0 Da); N-Y4: 4554.0 Da (theoretical value: 4556.1 Da); ^{15}N-N-Y4: 4614.0 Da (theoretical value: 4611.1 Da); N-Y5: 6053.7 (theoretical value: 6053.4); ^{15}N-N-Y5: 6119.5 Da (theoretical value: 6118.4 Da).

3.4.3 NMR Spectroscopy

For studies of structure or backbone dynamics 1 mM solution of the peptides at pH 5.6, 20 mM d_{13}-MES, 300 mM d_{38}-DPC were used. All spectra were recorded on an AV-700 Bruker NMR spectrometer at 310 K. Chemical shifts were calibrated to the water line at 4.63 ppm and nitrogen shifts were referenced indirectly to liquid NH_3\{Live et al., 1984, #80817\}. The spectra were processed using the Bruker Topspin2.0 software and transferred into the XEASY[35] or CARA[36] programs for further analysis.

For chemical shift assignments 3D ^{15}N-resolved TOCSY and NOESY[37] were used. In case of N-Y5 we decided to use ^{13}C,^{15}N labeling in combination with experiments that directly correlate sequential amide moieties[38]. In general experiments used coherence selection schemes via pulsed-field gradients[39] and sensitivity-enhancement building blocks[40] whenever possible. Upper-distance limits for structure calculations of N-Y1 were derived from a 70 ms NOESY spectrum[41]. Structures were calculated in the program CYANA[42], that uses restraint molecular dynamics in torsion angle space, and the implemented standard simulated annealing protocol in CYANA was used for that task.

A proton-detected version of the steady-state ^{15}N\{^{1}H\} heteronuclear Overhauser effect sequence were used for measurement of the hetereonuclear NOE[43]. Therein, the buildup of the NOE was achieved through a pulse train of 120 degree proton pulses separated by 5 ms over a period of 3 seconds.

For measurements of interactions by chemical shift mapping methodology 0.1 mM solutions of the ^{15}N-labeled neurohormones were mixed with the corresponding peptides from the N-terminal domains, and the deviations of peak positions were extracted from the [^{15}N,^{1}H]-HSQC spectra and computed according to Dd=SQRT(D(^{1}H)^2+0.2*D(15N)^2). Particular care was taken to ensure that no shifts in pH occurred when adding the N-Y peptides. In case of addition of various equivalents of pNPY to ^{15}N-labelled N-Y2 in the presence of DPC micelles (please describe the conditions here)

3.5 Acknowledgments

We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173)

3.6 References

1. Ma, P. & Zemmel, R. (2002). Value of novelty? *Nat Rev Drug Discov* **1**, 571-572.
2. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-45.
3. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-65.
4. Rasmussen, S. G. F., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., Schertler, G. F. X., Weis, W. I. & Kobilka, B. K. (2007). Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor. *Nature* **450**, 383-U4.
5. Murakami, M. & Kouyama, T. (2008). Crystal structure of squid rhodopsin. *Nature* **453**, 363-7.
6. Zou, C., Kumaran, S., Markovic, S., Walser, R. & Zerbe, O. (2008). Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family. *ChemBioChem* **9**, 2276-2284.
7. Larhammar, D. (1996). Structural diversity of receptors for neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **65**, 165-74.
8. Larhammar, D. & Salaneck, E. (2004). Molecular evolution of NPY receptor subtypes. *Neuropeptides* **38**, 141-51.
9. Fields, G. B. & Colowick, S. P. (1997). *Solid-Phase Peptide Synthesis*. Methods Enzymol., Academic Press, London.
10. Merrifield, B. (1997). Concept and early development of solid-phase peptide synthesis. *Methods Enzymol.* **289**, 3-13.
11. Sahdev, S., Khattar, S. & Saini, K. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* **307**, 249-264.
12. LaVallie, E. R., McCoy, J. M., Smith, D. B. & Riggs, P. (2001). Enzymatic and chemical cleavage of fusion proteins. *Curr Protoc Mol Biol* **16**, 4.9-4.17
13. Kuliopulos, A. & Walsh, C. (1994). Production, purification and cleavage of tandem repeats of recombinant peptides. *J Am Chem Soc* **116**, 4599-4607.
14. Bornstein, P. & Balian, G. (1977). Cleavage at Asn-Gly bonds with hydroxlyamine. *Meth Enzymol* **47**, 132-145.
15. Kohno, T., Kusunoki, H., Sato, K. & Wakamatsu, K. (1998). A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: an application to the production of mastoparan-X uniformly enriched with ^{15}N and $^{15}N/^{13}C$. *J Biomol NMR* **12**, 109-21.
16. Kapust, R. B., Tozser, J., Copeland, T. D. & Waugh, D. S. (2002). The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* **294**, 949-55.
17. Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D. & Waugh, D. S. (2001). Tobacco etch virus protease: mechanism of

autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**, 993-1000.
18. Mohanty, A., Simmons, C. & Wiener, M. (2003). Inhibition of tobacco etch virus protease activity by detergents. *Protein Expres Purif* **27**, 109-114-PII S1046-5928(02)00589-2.
19. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley.
20. Wishart, D., Sykes, B. & Richards, F. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* **222**, 311-33.
21. Palmer, A. G. (2001). NMR Probes of molecular dynamics: Overview and comparison with other techniques. *Ann Rev Biophys Biomol Struct* **30**, 129-155.
22. Brown, L. R., Bösch, C. & Wüthrich, K. (1981). Location and orientation relative to the micelle surface for glucagon in mixed micelles with dodecylphosphocholine: EPR and NMR studies. *Biochim Biophys Acta* **642**, 296-312.
23. Wüthrich, K., Billeter, M. & Braun, W. (1984). Polypeptide Secondary Structure Determination by Nuclear Magnetic Resonance Observation of Short Proton-Proton Distances. *J. Mol. Biol.* **180**, 715-740.
24. Zerbe, O., Neumoin, A., Mares, J., Walser, R. & Zou, C. (2006). Recognition of neurohormones of the NPY family by their receptors. *Journal of receptor and signal transduction research* **26**, 487-504.
25. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. (2001).Structure and dynamics of micelle-bound neuropeptide Y: comparison with unligated NPY and implications for receptor selection. *J Mol Biol* **305**, 307-392.
26. Bettio, A., Dinger, M. C. & Beck-Sickinger, A. G. (2002). The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold. *Protein Sci* **11**, 1834-44.
27. Cowley, D. J., Hoflack, J. M., Pelton, J. T. & Saudek, V. (1992). Structure of neuropeptide Y dimer in solution. *Eur J Biochem* **205**, 1099-106.
28. Sargent, D. F. & Schwyzer, R. (1986). Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc Natl Acad Sci U S A* **83**, 5774-8.
29. Schwyzer, R. (1987). Membrane-assisted molecular mechanism of neurokinin receptor subtype selection. *EMBO J* **6**, 2255-9.
30. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109-24.
31. Ridder, A. N., Morein, S., Stam, J. G., Kuhn, A., de Kruijff, B. & Killian, J. A. (2000). Analysis of the role of interfacial tryptophan residues in controlling the topology of membrane proteins. *Biochemistry* **39**, 6521-8.
32. Lerch, M., Mayrhofer, M. & Zerbe, O. (2004). Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J Mol Biol* **339**, 1153-68.
33. Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R. & Zerbe, O. (2007). Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution: Determinants of Stability of a Helical Hairpin. *Journal of the American Chemical Society* **129**, 8811-7.
34. Miroux, B. & Walker, J. E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-98.

35. Bartels, C., Xia, T.-h., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The program XEASY for computer-supported spectral analysis of biological macromolecules. *J Biomol NMR* **6**, 1-10.
36. Keller, R. (2004). *The Computer Aided Resonance Assignment*, CANTINA Verlag, Goldau.
37. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the Overlap Problem in the Assignment of ^{1H} NMR Spectra of Larger Proteins by Use of Three-Dimensional Heteronuclear 1H-15N Hartmann-Hahn Multiple-Quantum Coherence and Nuclear Overhauser-Multiple Quantum Coherence Spectroscopy: Application to Interleukin 1b. *Biochemistry* **28**, 6150-6156.
38. Weisemann, R., Ruterjans, H. & Bermel, W. (1993). 3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. *J Biomol NMR* **3**, 113-20.
39. Keeler, J., Clowes, R. T., Davis, A. L. & Laue, E. D. (1994). Pulsed-field gradients: theory and practice. *Methods Enzymol* **239**, 145-207.
40. Kay, L. E., Keifer, P. & Saarien, T. (1992). Pure absorption gradient enhanced heteronuclear single-quantum correlation sepctroscopy with improved sensitivity. *J Am Chem Soc* **114**, 10663-5.
41. Kumar, A., Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1-6.
42. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78.
43. Noggle, J. H. & Schirmer, R. E. (1971). *The Nuclear Overhauser Effect - Chemical Applications*, Academic Press, New York.

3.7 SUPPLEMENTARY MATERIALS

In tables S1-S4 chemical shifts were referenced to the water line taken at 4.63 ppm at 310K. The ^{15}N scale was derived indirectly by multiplying the frequency of 0 ppm for protons (the Bruker parameter SF) by 0.101329118. Chemical shifts have been deposited in the BMRB data base under deposition codes 80.8933262 (N-Y1), 80.6873033 (N-Y2) and 80.74817093 (N-5).

Table S1: Chemical Shifts of N-Y1 in the presence of DPC micelles

	H^N	H'^{α}	H^{β}	others
Met 1	-	-	-, -	γCH_2 -, -; εCH_3 -
Asn 2	-	-	-, -	δNH_2 -, -
Ser 3	8.52	4.42	-, -	γOH -
Thr 4	8.32	4.28	4.01	γCH_3 1.17; γOH -
Leu 5	8.23	4.03	1.86, 1.86	γH 1.17; δCH_3 0.79, 0.72
Phe 6	7.97	4.46	2.96, 3.18	δH 7.19; εH -, - ; ζH -
Ser 7	8.02	4.31	3.86, 3.86	γOH -
Gln 8	8.25	4.28	2.04, 2.04	γCH_2 2.33; εNH_2 7.44
Val 9	7.90	3.94	2.05	γCH_3 0.88
Glu 10	8.26	4.09	1.85, 1.88	γCH_2 2.33; εH -
Asn 11	8.29	4.51	2.69, 2.69	δNH_2 7.58, 6.86
His 12	8.32	4.57	3.05, 3.19	δ^1NH -; δ^2H 7.09; ε^1H -; ε^2NH -
Ser 13	8.18	4.40	3.79, 3.79	γOH -
Val 14	8.15	4.00	2.01	γCH_3 0.81
His 15	8.30	4.61	3.09, 3.09	δ^1NH -; δ^2H 7.05; ε^1H -; ε^2NH -
Ser 16	8.22	4.62	3.77, 3.80	γOH -
Asn 17	8.17	4.19	2.63, 2.63	δNH_2 7.50, 6.81
Phe 18	8.22	4.50	2.99, 3.10	δH 7.17; εH 7.25; ζH -
Ser 19	8.17	4.29	3.79, 3.79	γOH -
Glu 20	8.38	4.14	1.96, 1.96	γCH_2 2.22; εH -
Lys 21	8.15	4.18	1.71, 1.71	γCH_2 1.37; δCH_2 1.94, 2.02; εCH_2 2.74; ζNH_3^+ -
Asn 22	8.13	4.57	2.68, 2.68	δNH_2 7.49, 6.81
Ala 23	8.13	4.11	1.36	
Gln 24	8.13	4.15	2.02, 2.02	γCH_2 2.31; εNH_2 7.49
Leu 25	8.01	4.16	1.64, 1.64	γH 1.54; δCH_3 0.88, 0.81
Leu 26	7.83	4.17	1.57, 1.57	γH 1.46; δCH_3 0.83, 0.79
Ala 27	7.74	4.19	1.23	
Phe 28	7.97	4.55	3.13, 3.13	δH 7.20; εH 7.04 ; ζH -
Glu 29	8.33	4.20	1.87, 1.87	γCH_2 2.18; εH -
Asn 30	8.30	4.62	2.78, 2.78	δNH_2 7.58, 6.86
Asp 31	8.19	4.50	2.58, 2.58	δH -
Asp 32	8.07	4.29	2.78, 2.78	δH -
Cys 33	8.29	4.75	2.59, 2.65	γSH -
His 34	8.36	4.66	3.09, 3.09	δ^1NH -; δ^2H 7.13; ε^1H -; ε^2NH -
Leu 35	8.32	4.51	1.62, 1.62	γH 1.44; δCH_3 0.86
Pro 36		4.44	1.91, 1.96	γCH_2 2.19,2.22 ; δCH_2 3.78, 3.83
Leu 37	8.05	4.18	1.57, 1.57	γH -; δCH_3 0.87, 0.81
Ala 38	8.18	4.27	1.32	
Met 39	8.22	4.38	2.06	γCH_2 2.55, 2.66; εCH_3 1.96
Ile 40	7.38	4.01	1.79	γCH_2 1.09, 1.37; δCH_3 0.82

Table S2: Amide proton and ^{15}N chemical shifts of N-Y1

	N	HN
Met 1		
Asn 2		
Ser 3	116.26	8.52
Thr 4	116.44	8.36
Leu 5	122.7	8.27
Phe 6	115.8	7.99
Ser 7	114.84	8.04
Gln 8	120.6	8.25
Val 9	118.8	7.9
Glu 10	122.35	8.27
Asn 11	118.23	8.29
His 12	119.81	8.26
Ser 13	115.76	8.24
Val 14	120.68	8.18
His 15	120.62	8.38
Ser 16	116.22	8.2
Asn 17	119.23	8.2
Phe 18		
Ser 19	115.9	8.21
Glu 20	122.3	8.41
Lys 21	120.03	8.16
Asn 22	118.11	8.13
Ala 23	123.09	8.16
Gln 24	117.35	8.13
Leu 25	120.96	8.03
Leu 26	118.6	7.84
Ala 27	122.05	7.74
5Phe 28	117.83	7.98
Glu 29	120.35	8.34
Asn 30	118.61	8.31
Asp 31	120.03	8.31
Asp 32	118.07	8.08
Cys 33	120.08	8.47
His 34	119.74	8.39
Leu 35	122.68	8.39
Pro 36		
Leu 37	120.63	8.05
Ala 38	122.21	8.2
Met 39	118.15	8.23
Ile 40	122.86	7.38

Table S3: Amide proton and ^{15}N chemical shifts of N-Y2

	N	HN
Met 1		
Gly 2	112.1	8.59
Pro 3		
Ile 4	120.1	8.16
Gly 5	112.9	8.42
Ala 6	123.6	8.07
Glu 7	119.4	8.39
Ala 8	123.9	8.14
Asp 9	119.1	8.12
Glu 10	120.9	8.3
Asn 11	118.7	8.37
Gln 12	120.3	8.15
Thr 13	115.7	8.15
Val 14	122.2	8.09
Glu 15	123.8	8.33
Glu 16	121.5	8.26
Met 17	121.1	8.22
Lys 18	122.8	8.18
Val 19	121	8.02
Glu 20	123.9	8.37
Gln 21	120.6	8.15
Tyr 22	120.4	8.12
Gly 23	109.8	8.06
Pro 24		
Gln 25	119.8	8.49
Thr 26	114.7	8.08
Thr 27	118.5	8.09
Pro 28		
Arg 29	121	8.33
Gly 30	109.4	8.28
Glu 31	119.9	8.15
Leu 32	122.8	8.19
Val 33	122.6	8.04
Pro 34		
Asp 35	121.3	8.23
Pro 36		
Glu 37	121.8	8.3
Pro 38		
Glu 39	120.1	8.37
Leu 40	123.3	8.18
Ile 41	121.2	7.98
Asp 42	123.9	8.26
Ser 43	117.8	8.36
Thr 44	114.9	8.21
Lys 45	122.2	7.83
Leu 46	122.2	7.92
Ile 47	120.2	7.83

Glu 48	124.4	8.26
Val 49	121.4	8.06
Gln 50	128.5	7.89

Table S4: Amide proton and ^{15}N chemical shifts of N-Y4

Met 1		
Asn 2	8.171	119.838
Thr 3	8.29	114.787
Ser 4	8.26	127.348
His 5		
Leu 6		
Leu 7	8.006	121.818
Ala 8	7.953	123.247
Leu 9	7.817	119.79
Leu 10	7.915	121.832
Leu 11	7.897	123.675
Pro 12		
Lys 13	8.257	120.984
Ser 14	8.221	117.544
Pro 15		
Gln 16	8.338	119.671
Gly 17	8.243	109.457
Glu 18	8.225	120.207
Asn 19	8.454	119.731
Arg 20	8.252	121.321
Ser 21	8.211	116.37
Lys 22	8.092	123.461
Pro 23		
Leu 24	8.423	122.478
Gly 25	8.315	109.153
Thr 26	7.874	115.514
Pro 27		
Tyr 28		
Asn 29	8.083	120.162
Phe 30	7.989	120.843
Ser 31	8.069	116.045
Glu 32	8.133	121.685
His 33	8.288	117.461
Cys 34		
Gln 35	8.43	121.02
Asp 36	8.224	120.984
Ser 37	8.096	115.276
Val 38	8.002	120.362
Asp 39	8.251	122.983
Val 40	7.918	119.379
Met 41	7.85	128.423

Table S5: Amide proton and ^{15}N chemical shifts of N-Y5

	N	HN
Met 1		
Ser 2		
Phe 3	118.81	7.84
Tyr 4	118.03	7.63
Ser 5	115.85	8.1
Lys 6	122.27	8.25
Gln 7	119.04	8.18
Asp 8	119.37	8.03
Tyr 9	120.84	8.43
Asn 10	119.6	8.25
Met 11	120.23	8.22
Asp 12	120.96	8.38
Leu 13	121.09	7.9
Glu 14	120.56	8.39
Leu 15	121.51	8.39
Asp 16	118.01	8.47
Glu 17	118.35	8.04
Tyr 18	117.72	7.86
Tyr 19	118.5	8.1
Asn 20	117.52	8.21
Lys 21	118.77	7.9
Thr 22	113.27	7.8
Leu 23	121.69	7.83
Ala 24	122.63	7.76
Thr 25	112.49	7.92
Glu 26	122.74	8.19
Asn 27	120.23	8.43
Asn 28	119.29	8.35
Thr 29	114.05	8.05
Ala 30	125.61	8.17
Ala 31	122.26	8.07
Thr 32	112.43	7.92
Arg 33	122.34	8.32
Asn 34	119.25	8.36
Ser 35	115.51	8.17
Asp 36	121.69	8.26
Phe 37	120	8.11
Pro 38		
Val 39	117.15	8.04
Trp 40	122.72	8.12
Asp 41	118.58	8.12
Asp 42	118.49	7.95
Tyr 43	119.01	7.82
Lys 44	121.02	7.75
Ser 45	115.33	8.12
Ser 46	117.18	8.06
Val 47	118.82	7.84
Asp 48	122.02	8.04

Asp 49	119.12	8.03
Leu 50	121.18	7.9
Gln 51	124.45	7.67

Figure S6: Expansion of the 70 ms NOESY of N-Y1, displaying the assignment of residues 24 to 31:

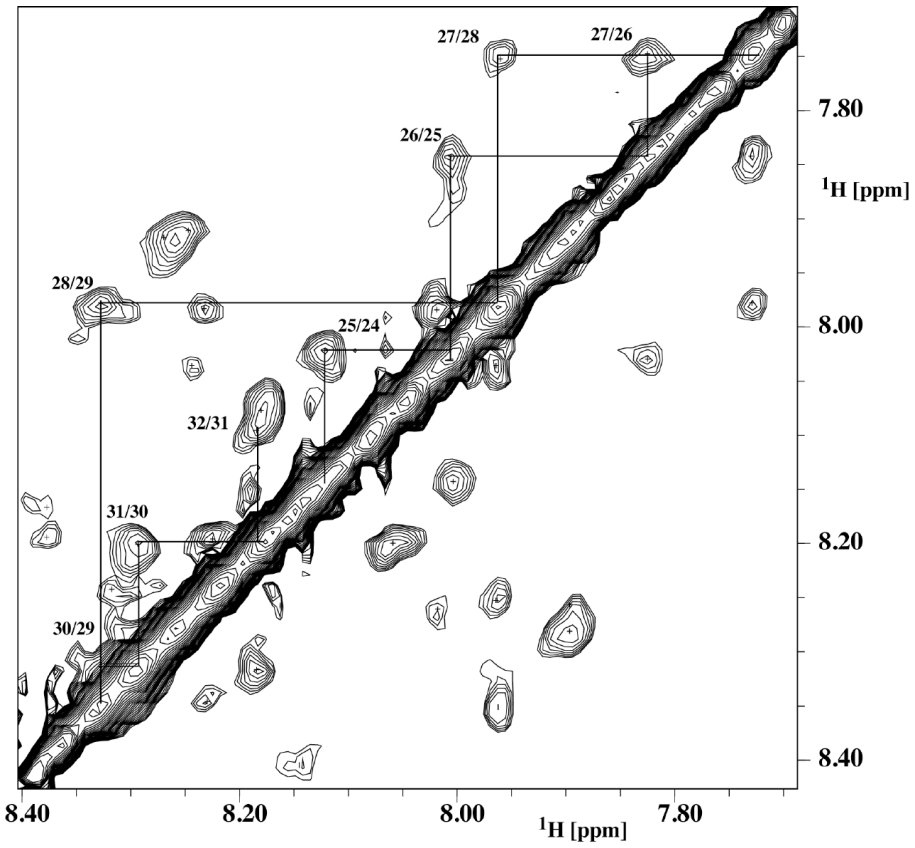
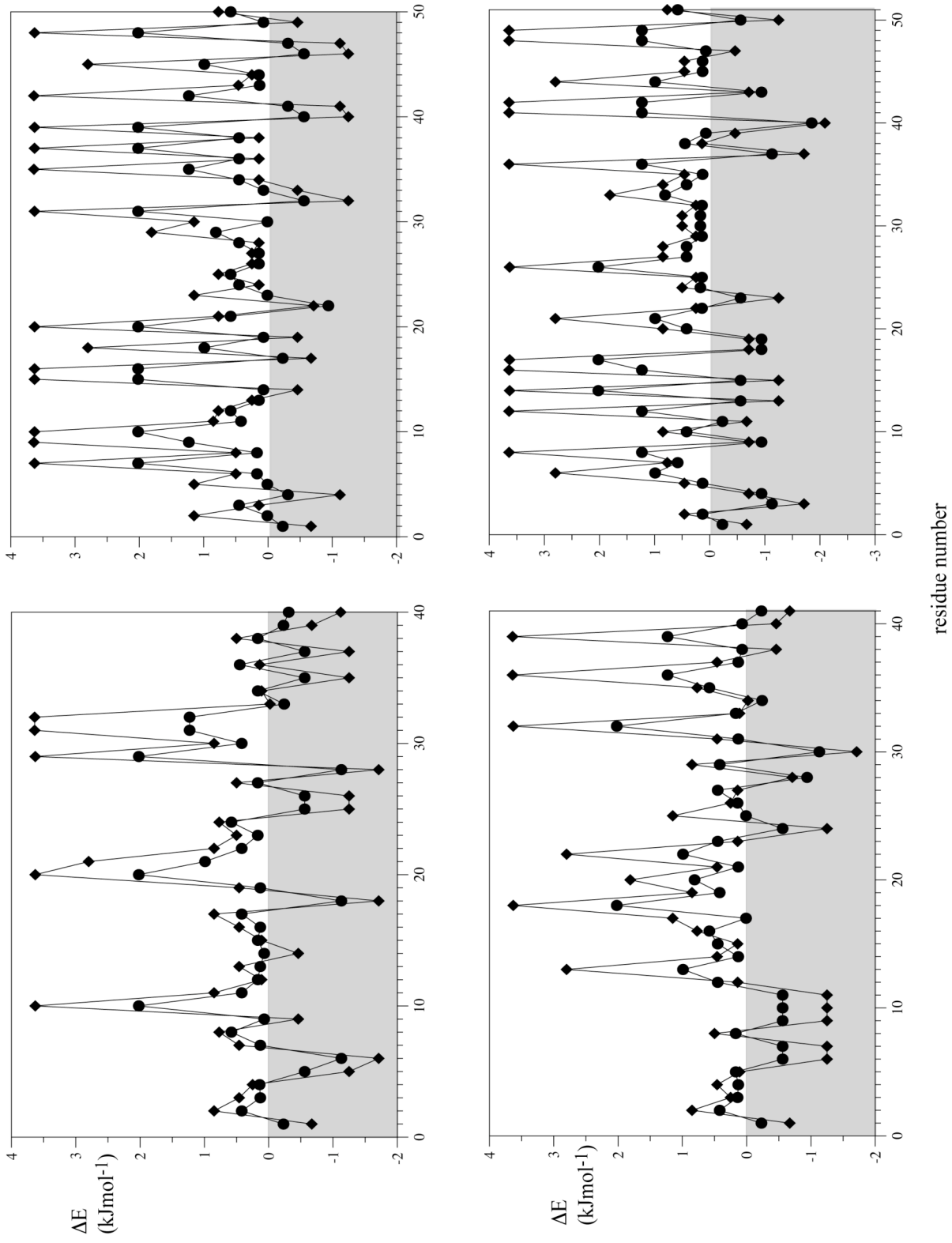


Figure S7: Free energies for partitioning residues of the Y receptor N-terminals (top left: N-Y1, top right: N-Y2, bottom left: N-Y4 and bottom right: N-Y5) domains into the water-membrane interface (circles) or into the membrane interior (diamonds) (data taken from ref. 24). Regions of favorable values are shaded in grey.



S8: Expression and purification of YUH

Yeast ubiquitin hydrolase (YUH) was expressed with a C-terminal hexahistidine tag and purified on a Ni-NTA column. The plasmid coding for the YUH-construct pYUHK20b was a generous gift from Toshiyuki Kohno (Mitsubishi Kasei Institute of Life Science, Tokyo, Japan).

5 ml of LB-broth containing 50 μg/ml kanamycin were inoculated with a colony of BL21 DE3 pYUHK20b cells, streaked onto plate from a glycerol stock, and incubated for 12 h at 37 ˚C and 220 rpm. 0.5 liter of LB-broth containing 50 μg/ml kanamycin were inoculated with the 5 ml overnight culture and incubated at 37 ˚C and 240 rpm in a 2 l Erlenmeyer flask. The culture was induced with 0.4 mM IPTG at an OD_{600} of 0.6-0.7 and grown for another 5 h to a final OD_{600} of around 5. The culture was harvested by centrifugation at 5000 rpm and 4 ˚C for 20 min. 7 g of wet biomass were obtained from 1 l of culture. The cell pellet was frozen at -20 ˚C.

The cell pellet was thawed on ice and resuspended in 40 ml resuspension buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol). 9 mg of lysozyme were added and the mixture incubated on ice for 15 min. The resuspension mixture was sonicated on ice with a Branson Digital Sonifier.

The lysate was centrifuged twice at 190000 rpm and 4 ˚C for 45 min and loaded onto a 10 ml column volume (CV) Ni-NTA-agarose column previously equilibrated with running buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 10 mM imidazole). Bound protein was eluted with elution buffer (50 mM Tris pH 8, 100 mM NaCl, 1 mM β-mercaptoethanol, 100 mM imidazole). The eluate was confirmed to contain the target protein by SDS-PAGE. To 10 ml of eluate 1.1 ml of glycerol were added to yield a final glycerol concentration of 10%. The YUH-concentration of this mixture was determined by a Bradford assay to be 7 mg/ml. This solution was stored at -20 °C in 1 ml aliquots.

S9: Expression and purification of TEV protease

The plasmid pTH24[1] was transformed to Rosetta(DE3) pLys cells. 0.5 ml of an overnight LB culture containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol were used to inoculate 1 liter of TB also containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. The culture was incubated at 37 °C. When it reached an OD_{600} of 0.6 it was induced with 0.1 mM IPTG and the temperature was lowered to 20 °C. After 20 hours the cells were harvested by centrifugation and the cell pellet was stored at –20 °C.

The cell pellet from 1 liter culture was resuspended in 40 ml washing buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 7). The resuspension mixture was sonicated on ice with a Branson Digital Sonifier. After centrifugation at 4 °C at 30000 g for 30 min the supernatant was loaded onto a Ni-NTA column. Unbound protein was eluted with washing buffer and bound protein was eluted with elution buffer (washing buffer with 200 mM imidazole). EDTA and DTT were added to a final concentration of 2 and 10 mM, respectively. 10 ml eluate were dialyzed over night against 1 liter dialysis buffer (25 mM sodium phosphate, 200 mM NaCl, 2 mM EDTA, 2 mM DTT, pH 8) at 4 °C. 10% glycerol were added and the protease solution was stored at –20 °C.

Susanne van den Berg, Per-Ake Lofdahl, Torleif Hard, Helena Berglund, Improved solubility of TEV protease by directed evolution, 2006, J. Biotechnol., 121, 291-298

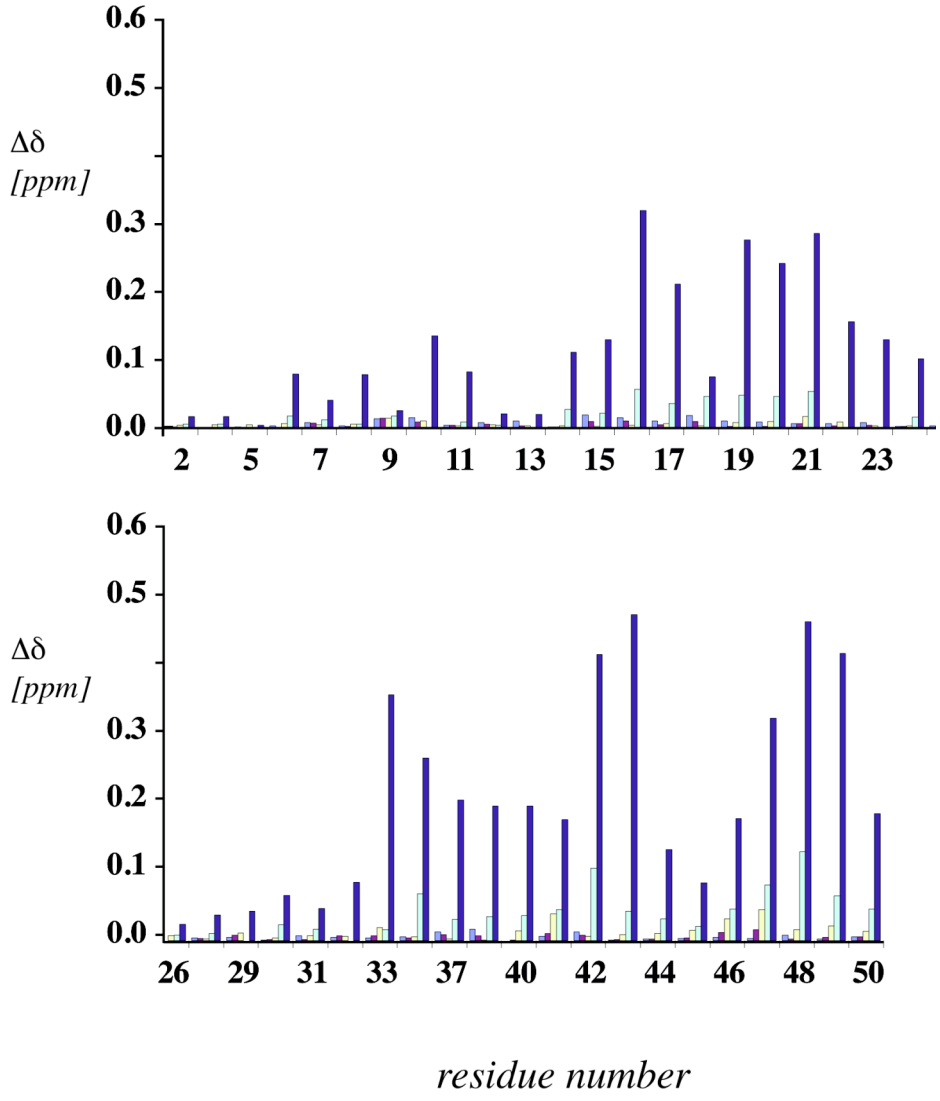
Table S10: MS fingerprinting of cleavage products from the N-Y4 expression:

Exp.	Theor.	_Mass	Peptide	fragment
2701	2701.279	0.278	(L)PKSPQGENRSKPLGTPYNFSEHCQ(D)	12-35
2701	2701.482	0.481	MNTSHLLALLLPKSPQGENRSKPLG(T)	1-25
2899	2899.582	0.582	MNTSHLLALLLPKSPQGENRSKPLGTP(Y)	1-27
3176	3176.688	0.688	MNTSHLLALLLPKSPQGENRSKPLGTPYN(F)	1-29
3323	3323.757	0.756	MNTSHLLALLLPKSPQGENRSKPLGTPYNF(S)	1-30
4554	4554.221	0.221	MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVM	1-41

Table S11: Screening of suitable detergents for enzymatic cleavage of the N-Y-KSI fusion peptides:

name	characteristics	concentration	100mMNaCl added
n-Decanoylsucrose	nonionic	9mM cannot solubilze the peptide	Improved a little
dihexanoylphosphatidylcholine	zwitterionic	4mM cannot solubilze the peptide	
n-Octyl-β-D-glucopyranoside	nonionic	35mM cannot solubilze the peptide	
n-Dodecyl-β-D maltopyranoside	nonionic	1-6x CMC cannot solubilze the peptide	
Fos-choline	zwitterionic	1-6x CMC cannot solubilze the peptide	Improved a little
n-Octyl-β-D-glucopyranoside	nonionic	1-6x CMC cannot solubilze the peptide	
n-Nonyl-β-D-maltoside	nonionic	1-6x CMC cannot solubilze the peptide	
n-Decyl-β-D-glucopyranoside	nonionic	1-6x CMC cannot solubilze the peptide	
Sarcosyl	anionic	0.4% can solubilze the peptide	completely soluble

Figure S12: Titration of N-Y2 with 0.5 (yellow), 1 (orange), 2 (red), 4 (purple), and 10 (blue) equivalents of pNPY.



Chapter 4

Biosynthesis and NMR-studies of a double transmembrane domain from the Y4 receptor, a human GPCR

4.1 Introduction

Membrane proteins are the most abundant class of proteins in prokaryotic and eukaryotic organisms and account for 20-30% of the total genome [1; 2]. Amongst these, G-protein coupled receptors (GPCRs) constitute the largest membrane protein family [3], accounting for 2% of the genome [4]. GPCRs play critical roles in molecular recognition and signal transduction and are among the most pursued pharmaceutical targets [5]. Around 30% of all marketed prescription drugs act on GPCRs, making this class of proteins a most successful therapeutic target [6].

Despite their prime biological importance surprisingly little structural information is available due to the tremendous difficulties encountered in producing GPCRs in active form and the problems associated with their structural study by crystallography or NMR. Recent advances in the expression and purification of membrane proteins have been described for various expression hosts, for example: *Escherichia coli* [7; 8; 9], yeast [10; 11], insect cells [12], mammalian cells [13; 14] and cell-free systems [15]. However, from approximately 1000 known GPCRs, only five high-resolution 3-D structures of two distinct receptor types have been reported: bovine rhodopsin [16] and opsin [17], squid rhodopsin [18], the human β2-adrenergic receptor [19; 20] and the turkey β1-adrenergic receptor [21].

As long as structural studies on intact GPCRs remain complicated by technical difficulties, the study of fragments of these receptors can deliver potentially valuable insights into the structure and function of these molecules. Studies on fragments may also help to establish methods required to tackle more complex systems, in particular by providing information concerning protein-lipid interactions. While fragments of domains from soluble proteins are often not stably folded, in integral membrane proteins the additional stabilizing interactions that occur between TM helices and the surrounding lipids can result in stretches of the polypeptide that are conformationally

defined and can be studied on their own. In 1990 Popot proposed a two-step model, the so-called partitioning-folding model, to describe assembly of membrane proteins in vivo [22; 23], that was later extended by White [24]: Initially, partitioning of the protein into the water-membrane interface results in formation of secondary structure. Interactions of the hydrophobic side chains with the surrounding lipid environment then lead to insertion of the transmembrane domains into the membrane interior. Finally, the functional protein is assembled via formation of the proper helix-helix contacts. According to this model the transmembrane domains can be thought of as independent folding units and be studied separately. A large body of literature supports the basic assumption of the model: For example, proteolysis of membrane proteins resulted in fragments containing entire TM sequences [25], and chemically or recombinantly synthesized TM peptides spontaneously assembled thereby rescuing receptor activity [26; 27; 28; 29]. Finally, peptides corresponding to the N and C terminus [30; 31], loop domains [32; 33; 34; 35] and transmembrane domains [33; 34; 36; 37; 38; 39; 40; 41; 42] from GPCRs have been found to fold to distinct secondary structures which in certain cases resembled the structures of the corresponding regions of the intact receptor.

TM domains usually contain about 25 residues [43; 44], therefore double-TM constructs in phospholipid micelles should be applicable to high-resolution NMR study. Though much effort has been devoted to the study of membrane proteins both by NMR and crystallography, so far few membrane protein structures have been determined by the former technique, amongst these the F_1,F_0-ATPase [45], the bacterial mercury transport membrane protein [46] and the human glycine receptor [47], all of which comprise two TM domains. One reason why there are still so few NMR studies of larger membrane proteins published is due to the fact that sufficient quantities of labeled protein are often not available for the required trials to optimize sample conditions. In the current study we therefore tried to optimize expression of a double transmembrane fragment of the NY-4 receptor. We consider that the solutions to problems addressed in this work might be generally applicable to researchers working on polytopic membrane polypeptides.

X-ray diffraction analysis of integral membrane proteins requires high quality single crystals. In contrast NMR in solution and the solid state is independent of protein crystallization and provides complementary information to that obtained by X-ray investigations [48; 49; 50; 51; 52; 53; 54]. However, NMR studies on GPCRs or large fragments of these integral membrane proteins require isotopic enrichment. This

requirement makes production impossible in expression systems such as mammalian hosts, because deuteration has not been achieved to date. Moreover, membrane proteins must be studied in a membrane-like environment such as detergent micelles. The concomitant increase in molecular weight as a consequence of micelle incorporation results in a dramatic decrease in spectral quality. In addition, slow conformational exchange processes lead to additional line-broadening. This has led to the frequently encountered experience that signals from the TM regions of membrane proteins remain invisible [54]. The lack of availability of fully deuterated detergents, compounds the technical difficulty of obtaining high resolution spectra for GPCRs or their fragments in micelles.

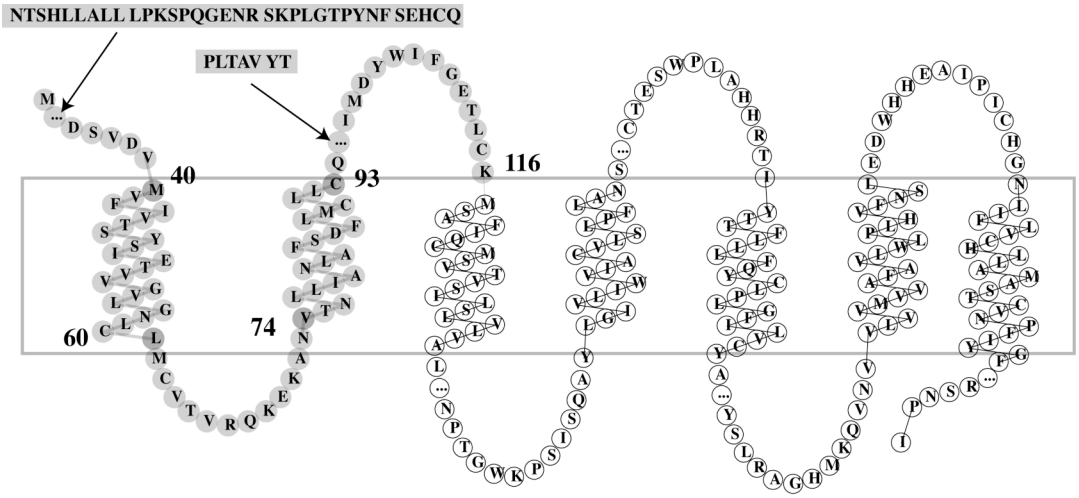


Figure 1: "Snake"-plot type presentation of the human Y4 receptor. The plot was modified from a download from the GPCR.org website. The part of the receptor that has been expressed in this work is shaded in gray. Note that the expressed polypeptide additionally contains a C-terminal $(His)_6$ tag. The omitted sequences for parts of the N terminus and the E1 loop are indicated separately in the figure.

Herein, we report on the expression and purification of a 115-residue (121 residues with the His tag) fragment from the neuropeptide Y4 GPCR containing the N terminus, the first transmembrane domain (T1), the first intracellular loop (I1), the second transmembrane domain (T2), and the first extracellular loop (E1) followed by a $(His)_6$ tag. This peptide (N-TM1-TM2) comprises about one third of the total length of the receptor (Fig. 1) and was obtained in multimilligram quantities. Importantly, the construct contains no fusion that needs to be removed after expression, and hence bypasses problems associated with chemical cleavage in the presence of residues like Met, or the enzymatic cleavage of hydrophobic sequences in the presence of detergents. Spectra with good quality could only be obtained when working under

reducing conditions which eliminated fragment oligomerization. Detergent mixtures proved to be necessary to yield the high quality spectra required for our analyses. Using a 1-palmitoyl-2-hydroxy-*sn*-glycero-3-[phospho-*rac*-(1-glycerol)] (LPPG)/dodecyl-phosphocholine (DPC) mixture and uniform $^{2}H,^{13}C,^{15}N$ labeling, TROSY-based 3D triple-resonance spectra could be recorded that allowed almost complete assignment of the backbone nuclei. The secondary chemical shifts indicate that the peptide is largely helical except for a mostly unfolded N-terminal domain.

4.2 Results

4.2.1 Optimization of Protein Expression

In order to obtain maximum expression of N-TM1-TM2 four different strains, BL21(DE3), C41(DE3) [55], BL21-AI and BL21-pLys(DE3), were evaluated. Amongst these BL21(DE3) is the most widely used expression host, while the other strains have been developed to express toxic proteins. Expression was tested for each strain at 37 °C and 20 °C. As shown in Fig. 2 temperature has a dramatic effect on the expression level of the target protein, which is significantly higher at 20 °C than at 37 °C. Although BL21(DE3) expresses the target protein at 20 °C, the reduced levels in comparison to the other strains that we tested indicates that the target protein may be toxic to this strain. Considering the perfect control of leakage expression, BL21-AI was chosen as the host for large-scale expression; nevertheless the difference in comparison to strains C41 or BL21pLys(DE3) is small.

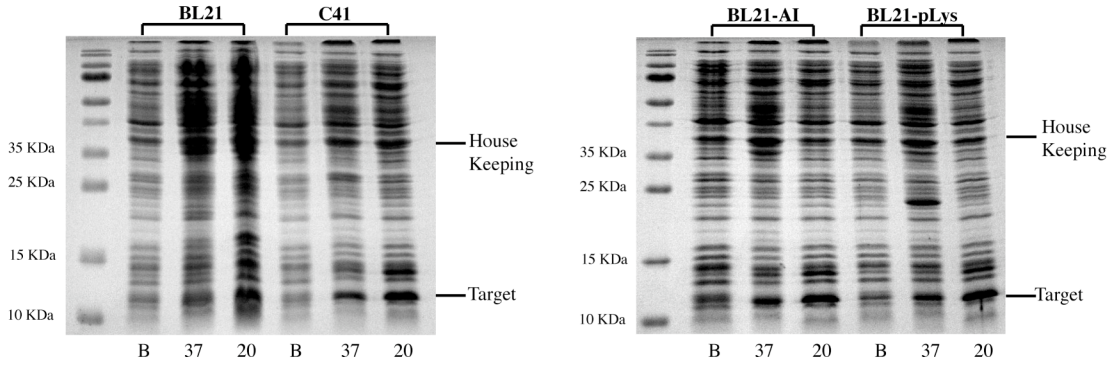


Figure 2: Selection of strain and expression conditions shown for BL21 and C41 (left) and for BL21-AI and BL21 pLys (right). B denotes "before induction", 37 denotes "induction at 37 °C" and 20 denotes "induction at 20 °C".

The chosen construct comprises six cysteine residues, some of which will spontaneously form disulfide bonds, in particular in the presence of the divalent cation Ni^{2+}. Protein preparations in both reducing and non-reducing sample buffer were analyzed by SDS-PAGE. It was observed that dimers, trimers and other oligomeric forms are observed in the non-reducing sample. Furthermore, we noticed the presence of a smear in the gel suggesting the occurrence of non-specific aggregation. Upon addition of 100 mM DTT to the sample buffer the smearing disappeared and the oligomerization was dramatically reduced indicating that disulfide bond formation was responsible for aggregation.

4.2.2 Optimization of Purification and Detergent

The protein recovered after Ni affinity chromatography and treatment with DTT was fairly homogeneous as judged by SDS-PAGE. Nevertheless, the [^{15}N,^{1}H]-TROSY spectrum still displayed too few peaks, and peak intensities varied considerably. The latter characteristic is most likely due to conformational exchange processes. We reasoned that lipid components from the cell membrane or other hydrophobic impurities that co-elute with N-TM1-TM2 from the affinity column may result in a conformationally heterogeneous interaction/integration into the phospholipid micelles. Using this protein preparation we were unable to identify detergents that resulted in better spectra (*vide infra*). Accordingly the eluant from the Ni affinity column was subjected to C4 reverse-phase HPLC. The detrimental effects on spectral quality of contaminants remaining after Ni affinity chromatography have also been recently discussed by Page *et al* [56]. The overall yield from a 1 L M9 culture of transformed BL-21AI cells after this additional step of chromatography was approximately 6 mg. We also noticed to our surprise that after lyophilization the solubility of the HPLC-purified protein in certain detergents had completely changed.

In order to obtain resolved TROSY spectra with sharp peaks a number of detergents were screened, including anionic (SDS, sarcosyl, LPPG, LMPG), zwitterionic (DPC, DHPC, LDAO) and non-ionic (OGP, DDM) detergents, and proton-nitrogen correlation spectroscopy was used to assess the suitability of the resulting samples for structural studies. As shown in Fig. 3 different detergents resulted in vastly different spectra. In some detergents tested the target protein was insoluble. Spectra measured in most detergents that dissolved the protein were of poor quality in that most of the expected peaks were missing and that some lines were very broad (Fig. 3G and H). Spectra recorded in the presence of SDS micelles resulted in too many peaks albeit that they were very sharp (Fig. 3F). In addition, measurements of the ^{15}N{^{1}H}-NOE indicated that the protein was highly flexible.

While the protein after elution from the Ni affinity column was nicely soluble in 200 mM LPPG solution, it turned out to be largely insoluble in the same detergent after the additional HPLC step. In contrast, it was now well soluble in DPC solution, a detergent in which the eluant from the Ni-affinity column was insoluble. Since it was observed that low-concentration samples prepared in LPPG resulted in good spectra,

and considering the fact that DPC can solublize the protein well, we tested mixtures of these two detergents to exploit the individual advantages of both. First the minimal

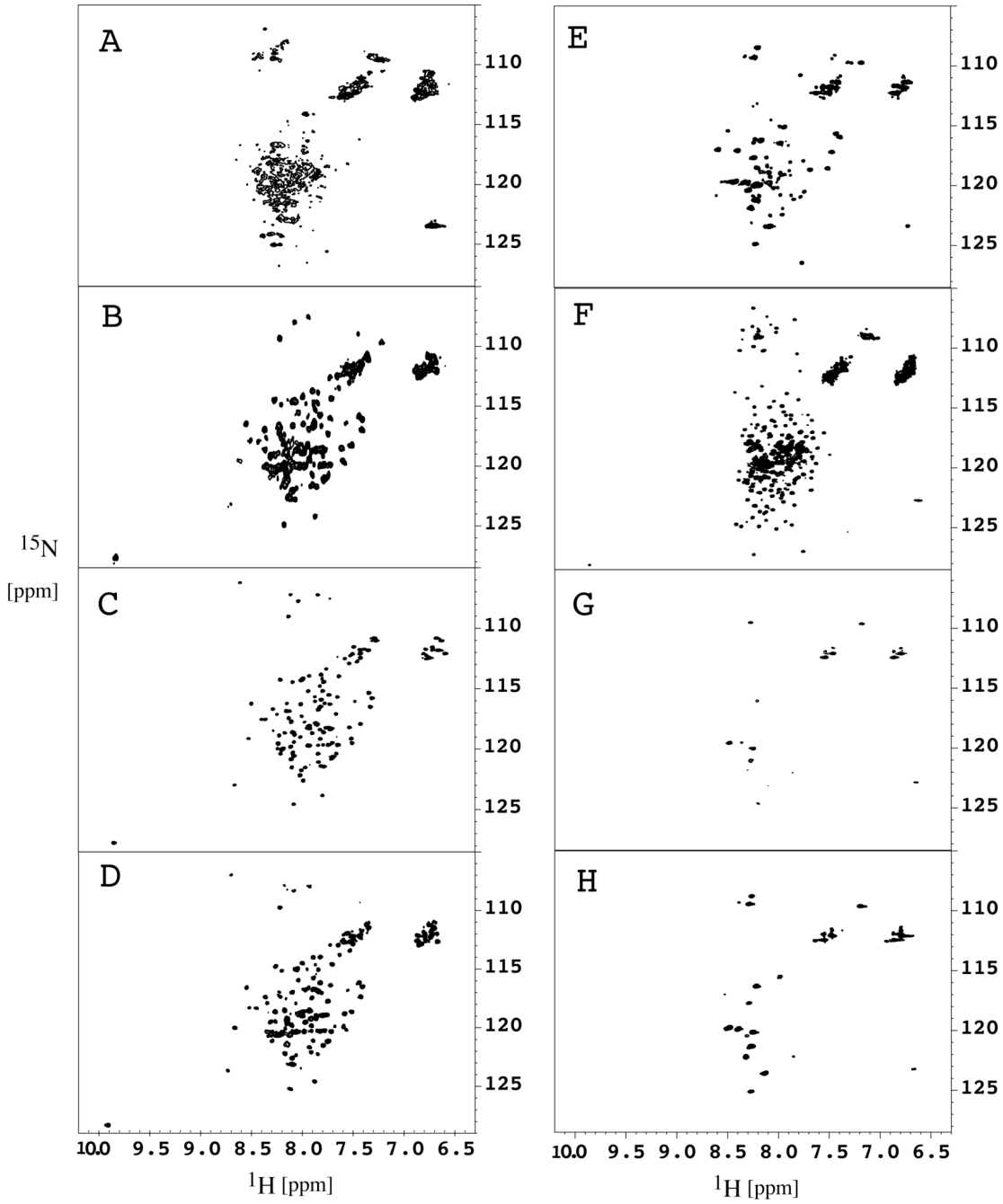


Figure 3: Plots of the two-dimensional [^{15}N,^{1}H]-HSQC spectra of N-TM1-TM2 recorded on samples of varying degrees of purity (spectra A to D) in various detergents (spectra E to H). Spectra were recorded using 0.3mM samples of the protein at pH 6.0 in 200mM LPPG (A,B,D), 30mM DPC/ 120mM LPPG (C), 150mM DPC (E), 170mM SDS (F), 170 mM OGP (G) and 100mM DHPC (H) at pH 6.0. The spectra on the left display protein samples directly after the Ni-affinity chromatography (A), after additional reduction with 100mM DTT and 250mM mercaptoethanol (B), after additional RP-HLPC in LPPG/DPC (C) and after purification and refolding using a method proposed by Page et al. [56] (D). The spectra on right were recorded with protein samples of highest purity and homogeneity. All data were recorded at 47 °C at 700 MHz proton frequency and the recognizable peak numbers out of the expected 115 are 74 (B), 109 (C), 97 (D), 63 (E), 161 (F), 12 (G), 15 (H), respectively, and is impossible to determine in (A).

concentration of DPC required to dissolve at least 0.5 mM protein was determined. Then increasing amounts of LPPG were added to DPC until a good-quality spectrum was obtained, and no further chemical shift changes upon addition of more LPPG occurred. The final detergent mixture consisted of 6% LPPG and 1% DPC and was used in all subsequent studies. The TROSY spectra recorded on such a sample displayed rather uniform linewidths. In addition, the $^{15}N\{^{1}H\}$-NOE data indicated that the backbone is rather rigid and that secondary structures are likely formed (see Fig. 6). Estimation of the overall correlation time derived from the ^{15}N R2/R1 ratio resulted in a value of 11.4 ns at 47°C.

4.2.3 Spectroscopy and Backbone Assignment

Considering the rather large molecular weight of the N-TM1-TM2/DPC/LPPG mixed micelle deuteration of the peptide was essential to yield spectra of sufficient quality. For backbone assignment a threefold strategy was pursued: i) matching of amide moieties via common Cab resonances in the HNCACB and HN(CO)CACB experiment, ii) matching via common CO frequencies in the HNCO and HN(CA)CO experiments, and iii) NOEs between sequential amide protons. Approx. 70% deuteration and the comparably narrow amide lines allowed for efficient TROSY-type triple resonance experiments. Alpha helical transmembrane proteins have intrinsically less signal dispersion and only constant-time ^{13}C and ^{15}N evolution in combination with mirror-image linear prediction provided sufficient resolution. Correlations in the triple-resonance HNCA and HNCACB spectra were observed for more than 80% of all residues. In the HNCO/HN(CA)CO pair correlations were almost always present. Representative strips from the assignment process are depicted in Fig. 4. Matching strips could be confirmed in the ^{15}N-resolved NOESY for all residues within the helical region with sufficient resolution in the proton frequency. In the end all H^N,N ,Ca and Cb nuclei could be assigned except for residues number 2 and 5, which are located in the flexible N terminal domain (see supplementary Table S3). Chemical shifts have been deposited in the BMRB database under accession code 15921.

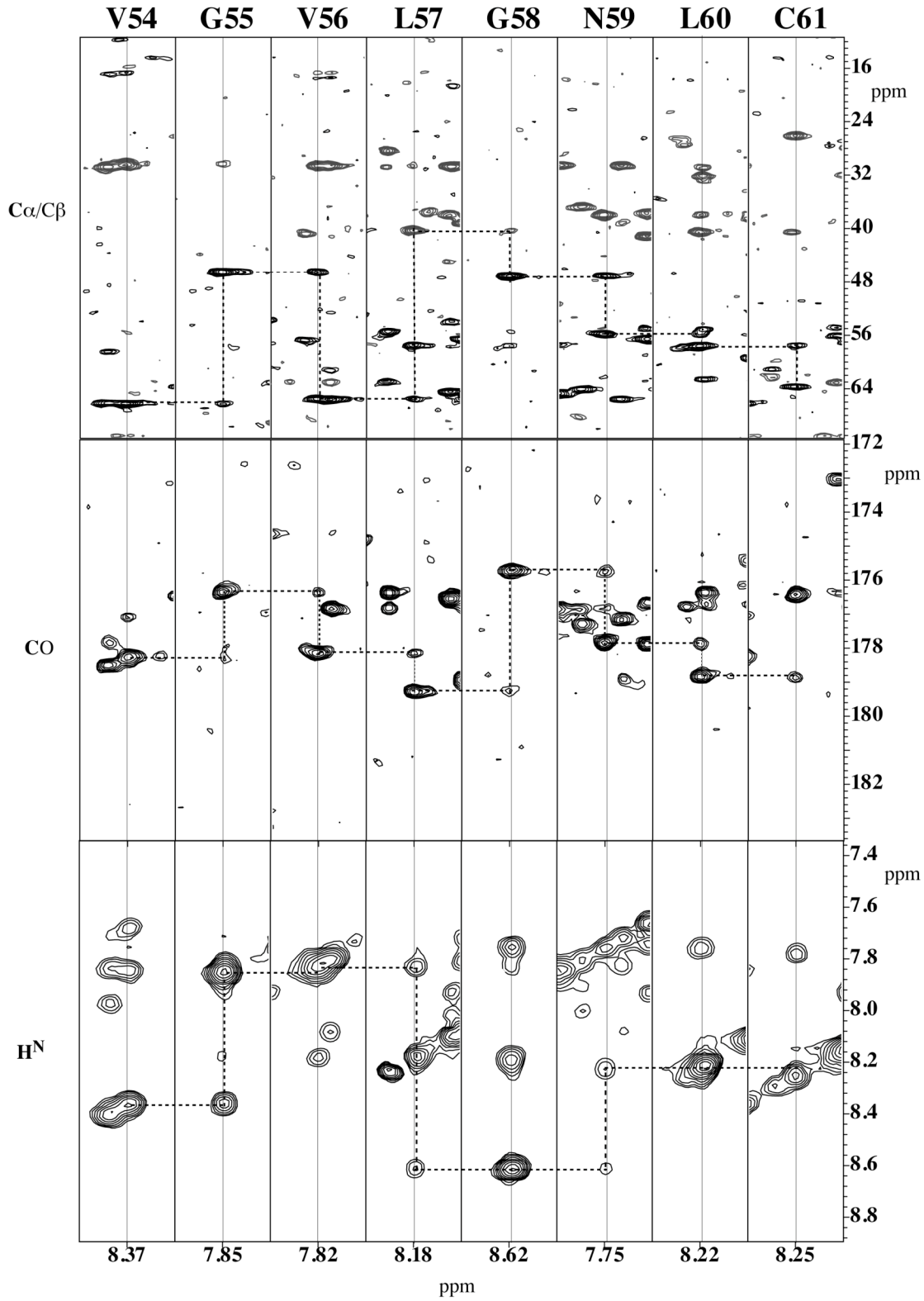


Figure 4: Plot displaying strips from the HNCACB (top), the HN(CA)CO (middle) and ^{15}N-NOESY spectra for the TM segment comprising residues Val54 to Cys61. Only Cα resonances are connected in the top panel. Strips were extracted at the ^{15}N chemical shifts of the corresponding amide nitrogen. All data were recorded at 700 MHz at 47 °C using the ^{2}H,^{13}C,^{15}N triply labeled protein in the 28 mM DPC/ 118 mM LPPG detergent mixture in 40 mM phosphate buffer, pH 6.0.

4.2.4 Secondary Structure

The CD spectrum of N-TM1-TM2 in DPC/LPPG mixed micelles is depicted in Fig. 5. For technical reasons, 50uM polypeptide was used in comparison to 0.5mM in the NMR sample. However based on the NMR spectra no aggregation occurred at the higher concentration and we believe the data obtained from the CD and NMR study is comparable. The CD spectrum clearly shows the presence of minima at 208 and 222 nm, typical for predominantly alpha helical conformations. In addition, deconvolution of the CD spectrum into contributions from the different secondary structural elements using the program K2D (http://www.embl-heidelberg.de/~andrade/k2d/) allowed estimating the content in a-helix to be around 57%. The CD analysis indicates that secondary structure under these conditions is properly formed.

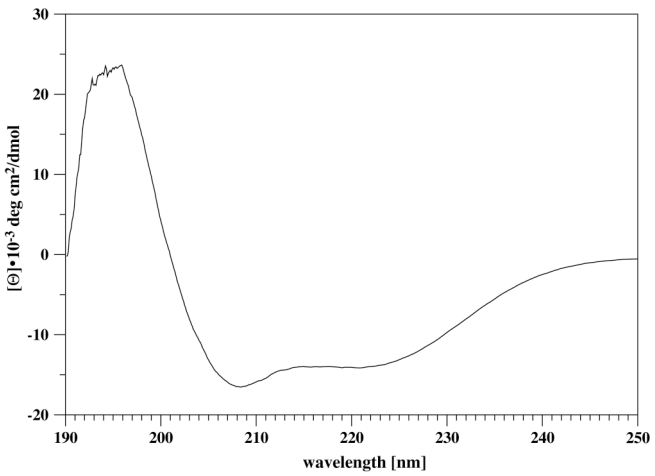


Figure 5: CD spectrum of 50 mM N-TM1-TM2 recorded at 47 °C in 40 mM phosphate buffer (pH 6.0) containing a mixture of 28 mM DPC and 118 mM LPPG. Data are converted to mean residue ellipticity.

In order to verify the results from the CD analysis, we have evaluated the $^{15}N\{^{1}H\}$-NOE to derive information on the rigidity at residue resolution. The data are depicted in Fig. 6 and compared to structural and dynamical properties of the isolated N-terminal domain from the Y4 receptor recently determined by us in the presence of pure DPC micelles at pH 5.6 [57]. The latter structural studies revealed the presence of a short a-helical stretch comprising residues 5 to 10, followed by a longer flexible loop in the segment between residues 11 and 25. Interestingly, the data on the construct described in this work indicated the presence of this flexible loop even when the N-terminal domain was fused to the first two helices. Otherwise the data indicate that with the exception of the N-terminal domain the protein is highly structured.

Surprisingly, little difference in rigidity is observed between residues from the putative TM helices and the loops. In addition the long first extracellular loop (E1), that in our construct lacks its native connection to the third TM, is rather rigid. Amide hydrogen exchange as measured in a [^{15}N,^{1}H]-HSQC experiment with and without presaturation of the water resonance revealed accelerated exchange only for the N-terminus, for the long unstructured loop in the N-terminal domain (see supplementary Figure S2) and in vicinity to the charged residue within TM1. Surprisingly, even in the I1 or E1 loop, hydrogen exchange is relatively slow indicating that these segments are reasonably folded and/or protected from solvent access.

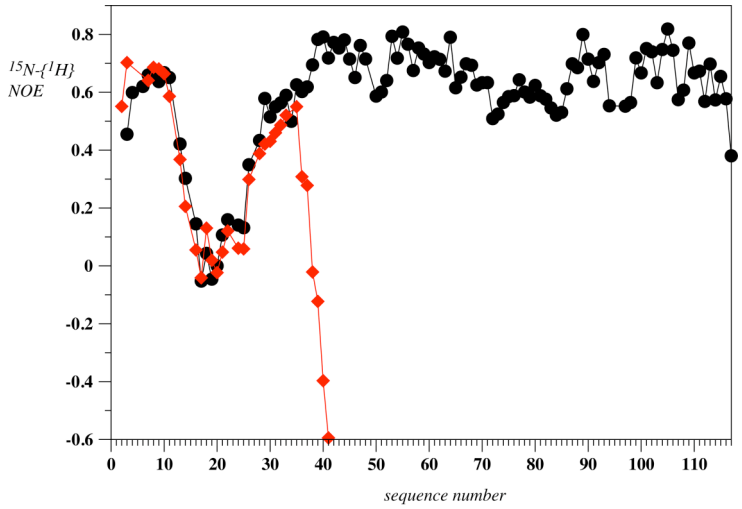


Figure 6: Comparison of the ^{15}N{^{1}H}-NOE values for N-TM1-TM2 (black spheres) described in this work and the isolated N-terminal domain from the Y4 receptor (N-Y4, red diamonds). All values were measured on the 600 MHz spectrometer. Data of N-Y4 are taken from Zou et al[57].

Sidechain assignment is presently in progress, which will help establishing secondary structure based on characteristic medium-range NOEs. However, backbone ^{15}N, Ca, Cb and C' shifts have already been assigned and hence the location and type of secondary structure can be predicted based on secondary chemical shifts [58; 59]. The output of the program TALOS [60] is depicted in Fig. 7. It predicts 74% of the 77 residue C-terminal fragment (the 2 TM helices plus the loops) to be helical. Interestingly, in both TM helices TALOS predictions indicate the TM helices to be destabilized adjacent to the internal polar residues Glu51and Thr52 in TM1 or Ser86 and Asp87 in TM2. Accordingly, no predictions were made for these regions. The locations of helical segments were also probed using proton,proton NOEs. In helices comparably short distances occur between sequential amide protons. Fig. 4 shows

contacts within the segment encompassing residues Val54 to Cys61 that are consistent with such short distances. Comparably strong NOEs between sequential amide protons occur through most of the residues in the TM1/TM2 segments. Additionally they are observed for most of the residues from the I1 and E1 loops.

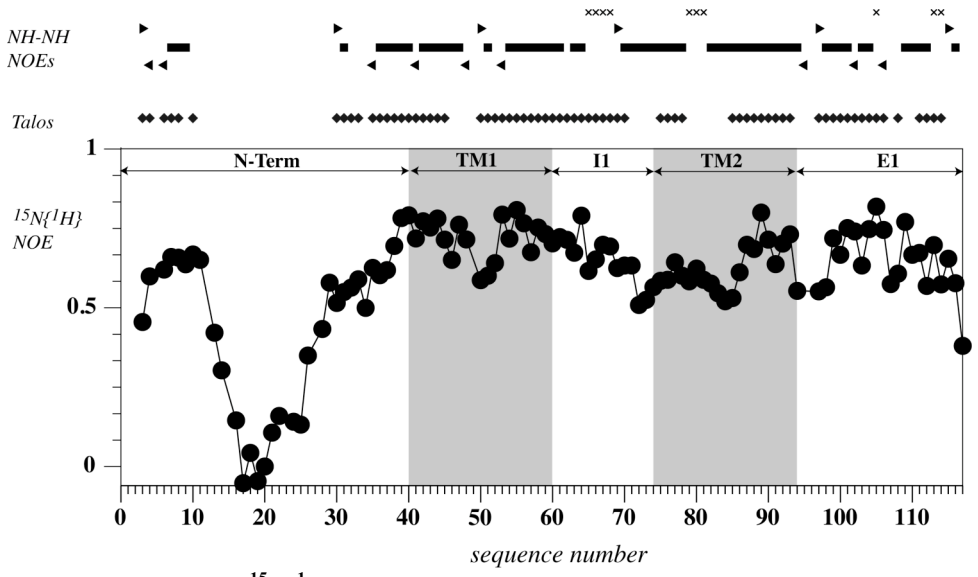


Figure 7: Summary of the $^{15}N\{^{1}H\}$-NOE values for N-TM1-TM2 (bottom), predicted regions of helical structure based on ^{15}N,$^{13}C\bar{\alpha}\bar{\beta}$, and C' chemical shifts using the program TALOS (middle) and the presence of NOEs between sequential amide protons (top). Amide moieties displaying NOEs to both preceding and following residues are indicated by squares, and by triangles with the top to the left or right for those residues that only display contact to predecessors or successor, respectively. All segments with degeneracy of proton chemical shifts that does not allow identification of NOE cross peaks are indicated by crosses.

4.3 Discussions

Considering the tremendous difficulties encountered during expression, purification, reconstitution and the spectroscopic evaluation of entire GPCRs, new strategies to derive useful structural information are highly desired. Accordingly, in this work we developed synthetic approaches for a double-TM construct that additionally contains the N-terminal domain and the first extracellular loop.

To our knowledge despite the success reported on the expression of polytopic bacterial membrane proteins[56], most multiple-TM polypeptides from higher organisms have been expressed as fusion proteins followed by either enzymatic or chemical cleavage from their fusion partners. Enzymes used to release the hydrophobic membrane peptides are often deactivated by the detergents that are required to solubilize the expressed fusion proteins. Thus yields are poor and much material is wasted. Cyanogen bromide (CNBr) is usually the chosen reagent for chemical cleavage, but is incompatible with the occurrence of internal methionine residues, limiting its general usage. In this study a relatively long double-TM domain (approx. one third of the sequence of the entire receptor) from a human receptor was expressed without a fusion partner. This approach allowed expression of the wild-type protein sequence, eliminated the cleavage step, simplified purification and resulted in a final yield of six mg/L of culture. It should be noted that expression of entire GPCRs has been accomplished in various hosts, as fusion proteins as well as directly, and work in this area has been reviewed [61; 62].

Purity and homogeneity are critical factors affecting the quality of NMR spectra. Considering that ^{15}N-NH_4Cl is comparably cheap and that [^{15}N,^{1}H]-TROSY spectra deliver a wealth of information on the state of the protein, we decided to monitor each step of purification using ^{15}N,^{1}H-correlation spectroscopy using only ^{15}N-labeled protein. We noticed a number of interesting points: (1) The Ni-NTA affinity chromatography seemed to result in pure protein as visualized by SDS-PAGE, however the spectral quality from such samples was clearly insufficient (see supplementary Figure S1); (2) due to the presence of 6 cysteines, the protein was prone to forming aggregates that result in severe line broadening, and work-up under strongly reducing conditions was mandatory (see supplementary Figure S1); (3) the dramatic improvement after HPLC purification indicated the presence of non-proteinaceous contaminants, which cannot be readily removed by affinity

chromatography. The chemical nature of the contaminants has not been identified so far, but we suspect them to be molecules that strongly associate with the protein so that they are not stripped off during the hydrophilic elution conditions of the affinity chromatography. This result suggests that they may be lipids or other hydrophobic components of the plasma membrane, that possibly also associate with the receptor in its natural environment. Another possibility is that they are proteins that bind to the metal affinity column. The presence of such contaminants apparently leads to heterogeneity in the microenvironment of the protein chains, in particular in the vicinity of the TM segments. This could affect the conformational exchange processes leading to the observed line-broadening. While HPLC purification is a standard technique for peptide chemists, it is often not used by protein biochemists because the solvent conditions denature most globular proteins. The possible presence of associating non-proteinaceous or proteinaceous contaminants is relevant to crystallographers who usually judge protein purity from SDS-PAGE gels. Perhaps screening of sample purity by $^{15}N,^{1}H$ NMR, at least for some of the smaller membrane proteins systems, could prove useful prior to embarking on crystallization attempts. We are aware that the proposed procedure requires a refolding step. In the context of entire GPCRs such refolding may not be achieved easily. However, in literature precedents that such refolding is possible can be found [63; 64; 65].

Membrane proteins can only properly exert their function when inserted in the membrane. Natural membranes, however, are characterized by the following features: they are patchy, with segregated regions of different chemical composition, variable thickness and distinct function [66]. To mimic this environment various media have been developed such as detergent micelles [67], bicelles [68; 69; 70] amphipols [71; 72], and very recently nanoscale bilayers [73] (for a general review on the usage of detergents in NMR studies of membrane proteins see [74; 75; 76]. For reasons of simplicity micelles have been frequently employed for NMR studies. In our study a wide range of detergents have been tested: Sarcosyl, LDAO, and DDM did not solubilize N-TM1-TM2. LPPG and LMPG only dissolved it to a very low extent, and others including DPC, OGP and DHPC dissolved the protein, but resulted in extremely broad spectra. Based on heteronuclear NOE analyses SDS resulted in a non-uniquely structured protein, an observation frequently also reported by other groups [67]. The result of the detergent screening conducted in this study indicated that it may be useful to consider detergent mixtures when optimizing membrane protein solubility and integration into

micelles. In the case of N-TM1-TM2 neither LPPG nor DPC gave satisfactory results, but the combination of these detergents resulted in a high-quality [^{15}N,^{1}H]-TROSY spectrum, in which 107 out of the expected 109 (without counting residues from the His-tag) peaks were observed. The final composition exhibited long-term stability and allowed us to run all of the three dimensional experiments required for a structural analysis. Natural membranes are heterogeneous mixtures of a variety of lipids and proteins. We suspect that various detergents can play different roles in solubilizing the peptide, aiding its integration into the lipid-like environment and forming a relatively stable composition. In the present example the LPPG head group is likely a much better mimic of head groups of naturally occurring lipids than DPC because the central glycerol component is retained. For reasons that are unclear to us at the moment, LPPG's capability to spontaneously allow insertion of the N-TM1-TM2 protein is low and it does not solubilize the purified polypeptide. In contrast DPC micelles readily integrate the membrane protein but give extremely broad lines in the HSQC spectra, possibly reflecting the presence of conformational exchange. The ratio between DPC and LPPG was, therefore, chosen to represent the minimal amount of DPC required to dissolve the protein. The optimized composition gave a highly resolved HSQC spectrum perhaps indicating that LPPG-peptide contacts are maximized in the TM region resulting in a relatively homogeneous microenvironment that led to good spectroscopic properties. By using a combination of detergents the number of membrane mimetic environments can be greatly increased and the possibility for trials that can exploit the synergistic contributions of different head groups and hydrophobic matches is maximized. It is important to note that protein detergent complexes are not idealized micelles and the insertion of detergents with different chain lengths at various positions in an asymmetric composition might, from a thermodynamic perspective, be predicted to lead to an optimally packed protein-lipid.

Inspection of NOEs between sequential amide protons, and restraints from chemical shifts delivered by TALOS allowed the derivation of the first low-resolution picture of secondary structure in the N-TM1-TM2 polypeptide. Stretches of the putative TM helices are predominantly helical (see Fig. 7). However, in the regions proximal to polar residues in the TMs (E and D in TM1 and TM2, respectively) the helices are destabilized, as judged by the reduction in the heteronuclear NOEs, by the TALOS predictions, by enhanced amide proton exchange and by the absence of

contacts between sequential amide protons. Buried glutamic acid and aspartic acid residues are rarely found in TM domains of integral membrane proteins, and we have noted such increased flexibility on another isolated TM domain in DPC micelles[42]. The biological significance of these findings will be subject to future work. A particularly interesting finding is, that the I1 and E1 loops are predominantly helical. The sequence of the beginning of the I1 loop is amphiphilic, and may possibly form a surface-associated helix. The sequence of the E1 loop is also amphiphilic in nature. In addition, it is rich in aromatic residues that are expected to position it in the interfacial compartment. Given the strong energetic driving force to place E1 in the interface compartment it is unlikely that E1 forms a flexible loop that diffuses into bulk solution. In the published crystal structures from rhodopsin [16] and the b-adrenergic receptors [20; 21], the long E2 loop contained elements of secondary structure; in the case of rhodopsin a short b-sheet, in the case of the β1- and β2-adrenergic receptors a-helices. However, the I1 and the E1 loops were devoid of regular secondary structure. Whether the helical nature of the E1 and I1 domains of N-TM1-TM2 is biologically relevant awaits additional studies on larger Y4 receptor fragments. At present it is also unclear how the I1 and E1 helices would connect the TM helices and reinsert smoothly into the membrane. However, in GPCR structures published to date we note that the length of the TM helices is not generally conserved, e.g. the TM5 and TM6 of squid rhodopsin were surprisingly deeply penetrating into the cytosol [18].

Previously, we reported the conformational preferences of the isolated N-terminal domain in the presence of DPC micelles [57]. The comparison of the dynamics data indicate that the latter and the corresponding fragment from the N-TM1-TM2 protein are highly similar in that they contain a short helix comprising residues 5 to 10, followed by a long and unstructured loop between residues 11 and 30. The segment that connects that loop to the first TM (residues 31 to 40) is rather flexible in the isolated N-Y4 peptide, but mostly helical in N-TM1-TM2. The amphiphilic sequence of the N-terminal region of N-TM1-TM2 is compatible with the presence of a surface-associated helix. Such a helix was also observed by us on a similar construct from the Ste2p receptor, a family D GPCR from yeast (unpublished results).

4.4 Conclusions

To conclude we have developed a synthetic route for directly expressing and isolating double-domain mammalian GPCR fragment in isotopically-labelled form in good yield. Rigorous purification using a combination of affinity chromatography and reversed-phase HPLC resulted in a sample with dramatically altered biophysical properties. A rational method for NMR sample optimization is introduced that relies on mixtures of detergents. The methodology allowed the collection of good-quality 3D NMR spectra, and preliminary results indicated the protein to be highly structured in the LPPG/DPC mixed micelles. Future work will be aimed at fully establishing the secondary and tertiary structure of this important domain of human N-Y4. We believe that the presented methodology may also be useful in the studies of even larger fragments or entire receptors.

4.5 Materials and methods

4.5.1 Plasmid Construction

The forward primer CGCGCT**CATATG**ATGAACACCTCTCACCTCCTG, in which bold letters denote a NdeI cleavage site and the backward primer AGCGCG**GGATCC***TCA*GTGATGGTGATGGTGATGCTTGCAGAGGGTCTCTC CAAA, in which bold letters denote a BamHI cleavage site, italic letters the stop codon and underlined letters the 6xHis tag, were used to amplify the gene encoding N-TM1-TM2 from the cDNA of the Y4 receptor (University of Missouri-Rolla, USA). The amplified gene was ligated into the plasmid pLC01 after both were cleaved with NdeI and BamHI and purified from agarose gel. The correctness of the recombinant DNA was confirmed by dideoxy sequencing (Synergene Biotech, Switzerland).

4.5.2 Protein Expression and Purification

The plasmid encoding the target protein was transformed into BL21-AI cells for expression, which were previously shown to result in higher expression levels compared to other strains [36]. A freshly transformed colony was used to inoculate 10 ml LB containing 100 mg/ml ampicillin. This preculture was grown over night at 37 °C and was then used to inoculate 1L LB (for the unlabeled sample) or M9 (with $^{15}NH_4Cl$ and ^{13}C glucose as sole nitrogen and carbon sources) media containing 100 mg/ml ampicillin and cultured at 37 °C until the OD_{600} reached 0.45-0.5. For induction the temperature was lowered to 20 °C and 0.2% L-arabinose was added. Cells were harvested after 12 hours and stored at –20 °C until further use. To allow expression in deuterated water transformed BL21-AI cells were plated on a D_2O M9 agar plate, and one colony was used to inoculate a LB preculture in 100% D_2O containing 100 mg/ml ampicillin. The preculture was grown at 37 °C overnight and was then used to inoculate 1L 95% D_2O M9 containing 75 mg/ml ampicillin. After incubation at 37°C overexpression was induced when the OD_{600} had reached 0.45 by adding 0.2% L-arabinose at 20 °C, and cells were harvested after 24 hours.

The cell pellet from 1 L culture was resuspended in GdHCl-containing buffer and the target protein purified from inclusion bodies under denaturing conditions using Ni-affinity chromatography. The protein was incubated together with 100 mM DTT, 250 mM mercaptoethanol, 10 mM EDTA at 4 °C over night to reduce any disulfide bonds.

The reduced eluant was purified by C4 reverse-phase HPLC using a H_2O/acetonitrile solvent system containing 0.1% TFA. The correctness of the target peptide was confirmed by MALDI-TOF (in case of unlabeled sample: 13645, theoretical mass: 13647.9) as well as western blotting with anti-His antibody and N-terminal amino acid sequencing. The level of deuteration for the sample that was used for the backbone assignment was approx. 65% according to MS. Incomplete deuteration is solely due to back-exchange from labile protons and protons picked up from the non-deuterated glucose.

4.5.3 NMR Sample Preparation

1.7 mg ^{15}N or $^{2}H,^{13}C,^{15}N$ uniformly labeled protein was dissolved in 200 µl 90%H_2O/ D_2O containing 2.5 mg DPC by thorough sonication and shaking at 37 °C for 30 min. 15 mg LPPG were dissolved in 50 µl 0.2 mM phosphate buffer (pH 6.0), after which the two detergent solutions were mixed. The final concentration for each component in the final solution was as follows: 0.5 mM protein, 1% (28 mM) DPC, 6% (118 mM) LPPG, 10% D_2O and 40 mM phosphate buffer (pH 6.0). The sample was stable for more than 2 months at 4 °C and more than 2 weeks at 47 °C.

4.5.4 NMR Spectroscopy and Backbone Assignment

All data were recorded on Avance 600 and 700 MHz Bruker spectrometers using triple-resonance cryoprobes at 47 °C. Chemical shifts of protons were calibrated according to the water line at 4.53 ppm at 47 °C, from which the carbon and nitrogen chemical shifts were referenced indirectly using the conversions factors published on the BMRB database. Sample optimization was conducted using solely ^{15}N-labeled samples and [$^{15}N,^{1}H$]-TROSY spectroscopy [77]. For backbone assignments standard Bruker experiments for the TROSY versions [78] of the 3D HNCACB [79; 80], HN(CO)CACB [79], HNCO [81] and HN(CA)CO [81] and a 200 ms ^{15}N-NOESY were used. For the HNCACB or HN(CO)CACB experiments 1024(^{1}H)*20(^{15}N)*80(^{13}C), for the HNCO or HN(CA)CO experiments 1024(^{1}H)*20(^{15}N)*32(^{13}C), and for the 3D ^{15}N-resolved NOESY 1024(^{1}H)*20(^{15}N)*125(^{1}H) complex data points were acquired. Spectral widths (and carrier positions) were 26 ppm (118.0 ppm) for ^{15}N, 60 ppm for ^{13}C in the experiments that label Ca and Cb resonances with the carbon carrier at 39 ppm for Cab and 54 ppm for Ca. In the HNCO-type experiments 20 ppm were used

for carbon, with the carrier set to 176 ppm. All experiments used pulsed field gradients for water suppression [82], and the Kay-Palmer sensitivity enhancement trick [83] as incorporated into the TROSY sequences by Weigelt [84]. A proton-detected version of the steady-state $^{15}N\{^{1}H\}$ heteronuclear Overhauser effect sequence was used for measurement of the heteronuclear NOE using a train of 120 degree proton pulses separated by 5 ms over a period of 3 seconds to achieve saturation of amide protons [85]. $^{15}N\{^{1}H\}$-NOEs were computed from the ratio of integrals from signals in the presence to those in the absence of amide proton irradiation.

Spectra were processed within the Bruker spectrometer software Topspin 2.0. Backbone assignment was accomplished within the software CARA [86]. Preferences for secondary structure based on ^{13}Ca, ^{13}Cb, ^{13}CO and ^{15}N chemical shifts were computed with the program TALOS [60].

4.5.5 Circular Dichroism Spectroscopy

CD spectra were recorded on Jasco model J-810 using 50 mM protein in 40 mM phosphate buffer (pH 6.0) in a mixture of 1% DPC and 6% LPPG in a quartz cuvette with a path length of 1 mm. All spectra were averaged from 3 consecutive measurements in the range between 190 and 250 nm at 47 $^{\circ}C$ with a slit width of 1nm and a scanning rate of 5 nm/min. The blank sample was recorded under identical conditions and subtracted from the sample spectra. The final CD intensity is expressed as the mean residue ellipticity (deg cm^2 $dmol^{-1}$).

4.6 ACKNOWLEDGEMENTS

We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173 to CZ), from the Alfred Werner Legat (to OZ) and from the National Institutes of Health (GM22086 to FN).

4.7 References

1. Boyd, D., Schierle, C. & Beckwith, J. (1998). How many membrane proteins are there? *Protein Sci* **7**, 201-5.
2. Stevens, T. J. & Arkin, I. T. (2000). Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* **39**, 417-20.
3. Foord, S. M. (2002). Receptor classification: post genome. *Curr Opin Pharmacol* **2**, 561-6.
4. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science* **291**, 1304-51.
5. Jacoby, E., Bouhelal, R., Gerspacher, M. & Seuwen, K. (2006). The 7 TM G-protein-coupled receptor target family. *ChemMedChem* **1**, 761-82.
6. Hopkins, A. L. & Groom, C. R. (2002). The druggable genome. *Nat Rev Drug Discov* **1**, 727-30.
7. Drew, D., Froderberg, L., Baars, L. & de Gier, J. W. (2003). Assembly and overexpression of membrane proteins in Escherichia coli. *Biochim Biophys Acta* **1610**, 3-10.
8. Drew, D., Slotboom, D. J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl-Beinker, N. M., Lambert, W., Lerch, M., Daley, D. O., Van Wijk, K. J., Hirst, J., Kunji, E. & De Gier, J. W. (2005). A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci* **14**, 2011-7.
9. Grisshammer, R., White, J. F., Trinh, L. B. & Shiloach, J. (2005). Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview. *J. Struct. Funct. Genomics* **6**, 159-63.
10. Wedekind, A., O'Malley, M. A., Niebauer, R. T. & Robinson, A. S. (2006). Optimization of the human adenosine A2a receptor yields in Saccharomyces cerevisiae. *Biotechnol Prog* **22**, 1249-55.
11. Lee, B. K., Jung, K. S., Son, C., Kim, H., Verberkmoes, N. C., Arshava, B., Naider, F. & Becker, J. M. (2007). Affinity purification and characterization of a G-protein coupled receptor, Saccharomyces cerevisiae Ste2p. *Prot Expr Pur* **56**, 62-71.

12. Massotte, D. (2003). G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim Biophys Acta* **1610**, 77-89.
13. Yin, D., Gavi, S., Shumay, E., Duell, K., Konopka, J. B., Malbon, C. C. & Wang, H. Y. (2005). Successful expression of a functional yeast G-protein-coupled receptor (Ste2) in mammalian cells. *Biochem Biophys Res Commun* **329**, 281-7.
14. Werner, K., Richter, C., Klein-Seetharaman, J. & Schwalbe, H. (2008). Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* **40**, 49-53.
15. Klammt, C., Schwarz, D., Eifler, N., Engel, A., Piehler, J., Haase, W., Hahn, S., Dotsch, V. & Bernhard, F. (2007). Cell-free production of G protein-coupled receptors for functional and structural studies. *J Struct Biol* **158**, 482-93.
16. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-45.
17. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. (2008). Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*.
18. Murakami, M. & Kouyama, T. (2008). Crystal structure of squid rhodopsin. *Nature* **453**, 363-7.
19. Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Kuhn, P., Weis, W. I., Kobilka, B. K. & Stevens, R. C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-65.
20. Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H. J., Yao, X. J., Weis, W. I., Stevens, R. C. & Kobilka, B. K. (2007). GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**, 1266-73.
21. Warne, T., Serrano-Vega, M., Baker, J., Moukhametzianov, R., Edwards, P., Henderson, R., Leslie, A., Tate, C. & Schertler, G. (2008). Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-91.
22. Popot, J. L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**, 4031-7.
23. Popot, J. L. & Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922.
24. White, S. H. & Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28**, 319-65.
25. Huang, K. S., Bayley, H., Liao, M. J., London, E. & Khorana, H. G. (1981). Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *J Biol Chem* **256**, 3802-9.
26. Kahn, T. W. & Engelman, D. M. (1992). Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. *Biochemistry* **31**, 6144-51.
27. Ridge, K. D., Lee, S. S. & Yao, L. L. (1995). In vivo assembly of rhodopsin from expressed polypeptide fragments. *Proc Natl Acad Sci U S A* **92**, 3204-8.

28. Martin, N. P., Leavitt, L. M., Sommers, C. M. & Dumont, M. E. (1999). Assembly of G protein-coupled receptors from fragments: identification of functional receptors with discontinuities in each of the loops connecting transmembrane segments. *Biochemistry* **38**, 682-95.
29. Wrubel, W., Stochaj, U. & Ehring, R. (1994). Construction and in vivo analysis of new split lactose permeases. *FEBS Lett* **349**, 433-8.
30. Harmar, A. J. (2001). Family-B G-protein-coupled receptors. *Genome Biol* **2**, REVIEWS3013.
31. O'Hara, P. J., Sheppard, P. O., Thogersen, H., Venezia, D., Haldeman, B. A., McGrane, V., Houamed, K. M., Thomsen, C., Gilbert, T. L. & Mulvihill, E. R. (1993). The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41-52.
32. Bennett, M., Yeagle, J. A., Maciejewski, M., Ocampo, J. & Yeagle, P. L. (2004). Stability of loops in the structure of lactose permease. *Biochemistry* **43**, 12829-37.
33. Katragadda, M., Alderfer, J. L. & Yeagle, P. L. (2001). Assembly of a polytopic membrane protein structure from the solution structures of overlapping peptide fragments of bacteriorhodopsin. *Biophys J* **81**, 1029-36.
34. Katragadda, M., Chopra, A., Bennett, M., Alderfer, J. L., Yeagle, P. L. & Albert, A. D. (2001). Structures of the transmembrane helices of the G-protein coupled receptor, rhodopsin. *J Pept Res* **58**, 79-89.
35. Yeagle, P. L., Salloum, A., Chopra, A., Bhawsar, N., Ali, L., Kuzmanovski, G., Alderfer, J. L. & Albert, A. D. (2000). Structures of the intradiskal loops and amino terminus of the G-protein receptor, rhodopsin. *J Pept Res* **55**, 455-65.
36. Cohen, L. S., Arshava, B., Estephan, R., Englander, J., Kim, H., Hauser, M., Zerbe, O., Ceruso, M., Becker, J. M. & Naider, F. (2008). Expression and biophysical analysis of two double-transmembrane domain-containing fragments from a yeast G protein-coupled receptor. *Biopolymers* **90**, 117-30.
37. Zheng, H., Zhao, J., Sheng, W. & Xie, X. Q. (2006). A transmembrane helix-bundle from G-protein coupled receptor CB2: biosynthesis, purification, and NMR characterization. *Biopolymers* **83**, 46-61.
38. Musial-Siwek, M., Kendall, D. A. & Yeagle, P. L. (2008). Solution NMR of signal peptidase, a membrane protein. *Biochim Biophys Acta* **1778**, 937-44.
39. Tian, C., Vanoye, C. G., Kang, C., Welch, R. C., Kim, H. J., George, A. L., Jr. & Sanders, C. R. (2007). Preparation, functional characterization, and NMR studies of human KCNE1, a voltage-gated potassium channel accessory subunit associated with deafness and long QT syndrome. *Biochemistry* **46**, 11459-72.
40. Lau, T. L., Partridge, A. W., Ginsberg, M. H. & Ulmer, T. S. (2008). Structure of the integrin beta3 transmembrane segment in phospholipid bicelles and detergent micelles. *Biochemistry* **47**, 4008-16.
41. Mobley, C. K., Myers, J. K., Hadziselimovic, A., Ellis, C. D. & Sanders, C. R. (2007). Purification and initiation of structural characterization of human peripheral myelin protein 22, an integral membrane protein linked to peripheral neuropathies. *Biochemistry* **46**, 11185-95.
42. Neumoin, A., Arshava, B., Becker, J., Zerbe, O. & Naider, F. (2007). NMR studies in dodecylphosphocholine of a fragment containing the seventh transmembrane helix of a G-protein-coupled receptor from Saccharomyces cerevisiae. *Biophys J* **93**, 467-82.

43. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. & von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-81.
44. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H. & von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-30.
45. Rastogi, V. K. & Girvin, M. E. (1999). Structural changes linked to proton translocation by subunit c of the ATP synthase. *Nature* **402**, 263-8.
46. Howell, S. C., Mesleh, M. F. & Opella, S. J. (2005). NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury detoxification system. *Biochemistry* **44**, 5196-206.
47. Ma, D., Liu, Z., Li, L., Tang, P. & Xu, Y. (2005). Structure and dynamics of the second and third transmembrane domains of human glycine receptor. *Biochemistry* **44**, 8790-800.
48. Mackenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A Transmembrane Helix Dimer - Structure and Implications. *Science* **276**, 131-133.
49. Getmanova, E., Patel, A. B., Klein-Seetharaman, J., Loewen, M. C., Reeves, P. J., Friedman, N., Sheves, M., Smith, S. O. & Khorana, H. G. (2004). NMR spectroscopy of phosphorylated wild-type rhodopsin: mobility of the phosphorylated C-terminus of rhodopsin in the dark and upon light activation. *Biochemistry* **43**, 1126-33.
50. Klein-Seetharaman, J., Reeves, P. J., Loewen, M. C., Getmanova, E. V., Chung, J., Schwalbe, H., Wright, P. E. & Khorana, H. G. (2002). Solution NMR spectroscopy of [alpha -15N]lysine-labeled rhodopsin: The single peak observed in both conventional and TROSY-type HSQC spectra is ascribed to Lys-339 in the carboxyl-terminal peptide sequence. *Proc Natl Acad Sci U S A* **99**, 3452-7.
51. Klein-Seetharaman, J., Yanamala, N. V., Javeed, F., Reeves, P. J., Getmanova, E. V., Loewen, M. C., Schwalbe, H. & Khorana, H. G. (2004). Differential dynamics in the G protein-coupled receptor rhodopsin revealed by solution NMR. *Proc Natl Acad Sci U S A* **101**, 3409-13.
52. Schubert, M., Kolbe, M., Kessler, B., Oesterhelt, D. & Schmieder, P. (2002). Heteronuclear multidimensional NMR spectroscopy of solubilized membrane proteins: resonance assignment of native bacteriorhodopsin. *ChemBioChem* **3**, 1019-23.
53. Oxenoid, K., Kim, H. J., Jacob, J., Sonnichsen, F. D. & Sanders, C. R. (2004). NMR assignments for a helical 40 kDa membrane protein. *J Am Chem Soc* **126**, 5048-9.
54. Tian, C., Breyer, R. M., Kim, H. J., Karra, M. D., Friedman, D. B., Karpay, A. & Sanders, C. R. (2005). Solution NMR spectroscopy of the human vasopressin V2 receptor, a G protein-coupled receptor. *J Am Chem Soc* **127**, 8010-1.
55. Miroux, B. & Walker, J. E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289-98.

56. Page, R. C., Moore, J. D., Nguyen, H. B., Sharma, M., Chase, R., Gao, F. P., Mobley, C. K., Sanders, C. R., Ma, L., Sönnichsen, F. D., Lee, S., Howell, S. C., Opella, S. J. & Cross, T. A. (2006). Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *Journal of structural and functional genomics* **7**, 51-64.
57. Zou, C., Kumaran, S., Markovic, S., Walser, R. & Zerbe, O. (2008). Studies of the structure of the N-terminal domain from the Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from the NPY family. *ChemBioChem* **9**, 2276-2284.
58. Wishart, D. S. & Sykes, B. D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**, 171-80.
59. Wishart, D., Sykes, B. & Richards, F. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* **222**, 311-33.
60. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302.
61. Sarramegn, V., Muller, I., Milon, A. & Talmont, F. (2006). Recombinant G protein-coupled receptors from expression to renaturation: a challenge towards structure. *Cell Mol Life Sci* **63**, 1149-64.
62. Sarramegna, V., Talmont, F., Demange, P. & Milon, A. (2003). Heterologous expression of G-protein-coupled receptors: comparison of expression systems fron the standpoint of large-scale production and purification. *Cell Mol Life Sci* **60**, 1529-46.
63. Baneres, J. L., Martin, A., Hullot, P., Girard, J. P., Rossi, J. C. & Parello, J. (2003). Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B4 binding to recombinant BLT1. *J Mol Biol* **329**, 801-14.
64. Kiefer, H., Krieger, J., Olszewski, J. D., Von Heijne, G., Prestwich, G. D. & Breer, H. (1996). Expression of an olfactory receptor in Escherichia coli: purification, reconstitution, and ligand binding. *Biochemistry* **35**, 16077-84.
65. Banères, J. L., Mesnier, D., Martin, A., Joubert, L., Dumuis, A. & Bockaert, J. (2005). Molecular characterization of a purified 5-HT4 receptor: a structural basis for drug efficacy. *J Biol Chem* **280**, 20253-60.
66. Engelman, D. M. (2005). Membranes are more mosaic than fluid. *Nature* **438**, 578-80.
67. Krüger-Koplin, R. D., Sorgen, P. L., Krüger-Koplin, S. T., Rivera-Torres, I. O., Cahill, S. M., Hicks, D. B., Grinius, L., Krulwich, T. A. & Girvin, M. E. (2004). An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR* **28**, 43-57.
68. Glover, K. J., Whiles, J. A., Wu, G., Yu, N., Deems, R., Struppe, J. O., Stark, R. E., Komives, E. A. & Vold, R. R. (2001). Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys J* **81**, 2163-71.
69. Vold, R. R., Prosser, R. S. & Deese, A. J. (1997). Isotropic Solutions Of Phospholipid Bicelles - a new membrane mimetic for high-resolution nmr studies of polypeptides. *J Biomol NMR* **9**, 329-335.

70. Poget, S. F. & Girvin, M. E. (2007). Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochim Biophys Acta* **1768**, 3098-106.
71. Gohon, Y., Dahmane, T., Ruigrok, R. W., Schuck, P., Charvolin, D., Rappaport, F., Timmins, P., Engelman, D. M., Tribet, C., Popot, J. L. & Ebel, C. (2008). Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophys J* **94**, 3523-37.
72. Zoonens, M., Catoire, L. J., Giusti, F. & Popot, J. L. (2005). NMR study of a membrane protein in detergent-free aqueous solution. *Proc Natl Acad Sci U S A* **102**, 8893-8.
73. Lyukmanova, E. N., Shenkarev, Z. O., Paramonov, A. S., Sobol, A. G., Ovchinnikova, T. V., Chupin, V. V., Kirpichnikov, M. P., Blommers, M. J. & Arseniev, A. S. (2008). Lipid-protein nanoscale bilayers: a versatile medium for NMR investigations of membrane proteins and membrane-active peptides. *J Am Chem Soc* **130**, 2140-1.
74. Sanders, C., Kuhn Hoffmann, A., Gray, D., Keyes, M. & Ellis, C. (2004). French swimwear for membrane proteins. *ChemBioChem* **5**, 423-6.
75. Sanders, C. R. & Oxenoid, K. (2000). Customizing model membranes and samples for NMR spectroscopic studies of complex membrane proteins. *Biochim Biophys Acta* **1508**, 129-45.
76. Sanders, C. R. & Sönnichsen, F. (2006). Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* **44**, S24-40.
77. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* **94**, 12366-71.
78. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. (1999). TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J Am Chem Soc* **121**, 844-848.
79. Shan, X., Gardner, K., Muhandiram, D., Rao, N., Arrowsmith, C. & Kay, L. (1996). Assignment of N-15, C-13(alpha), C-13(beta), and HN resonances in an N-15, C-13, H-2 labeled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and H-2-decoupling. *J Am Chem Soc* **118**, 6570-6579.
80. Wittekind, M. & Mueller, L. (1993). HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *J Magn Reson Ser B* **101**, 201-205.
81. Yamazaki, T., Lee, W., Arrowsmith, C., Muhandiram, D. & Kay, L. (1994). A suite of triple-resonance NMR experiments for the backbone assignment of N-15,C-13,H2- labeled proteins with high sensitivity *J Am Chem Soc* **116**, 11655-11666.
82. Keeler, J., Clowes, R. T., Davis, A. L. & Laue, E. D. (1994). Pulsed-field gradients: theory and practice. *Methods Enzymol* **239**, 145-207.
83. Kay, L. E., Keifer, P. & Saarien, T. (1992). Pure absorption gradient enhanced heteronuclear single-quantum correlation sepctroscopy with improved sensitivity. *J Am Chem Soc* **114**, 10663-10665.
84. Weigelt, J. (1998). Single scan, sensitivity- and gradient-enhanced TROSY for multidimensional NMR experiments. *Journal of the American Chemical Society* **120**, 10778-10779.

85. Noggle, J. H. & Schirmer, R. E. (1971). *The Nuclear Overhauser Effect - Chemical Applications*, Academic Press, New York.
86. Keller, R. (2004). *The Computer Aided Resonance Assignment*, CANTINA Verlag, Goldau.

4.8 Supplementary Materials

Figures: S1: Aggregation for TM1-TM2

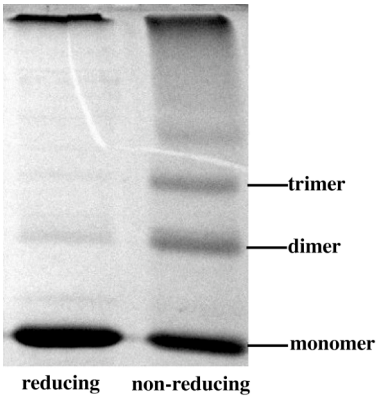


Figure S2: Presaturation experiment:

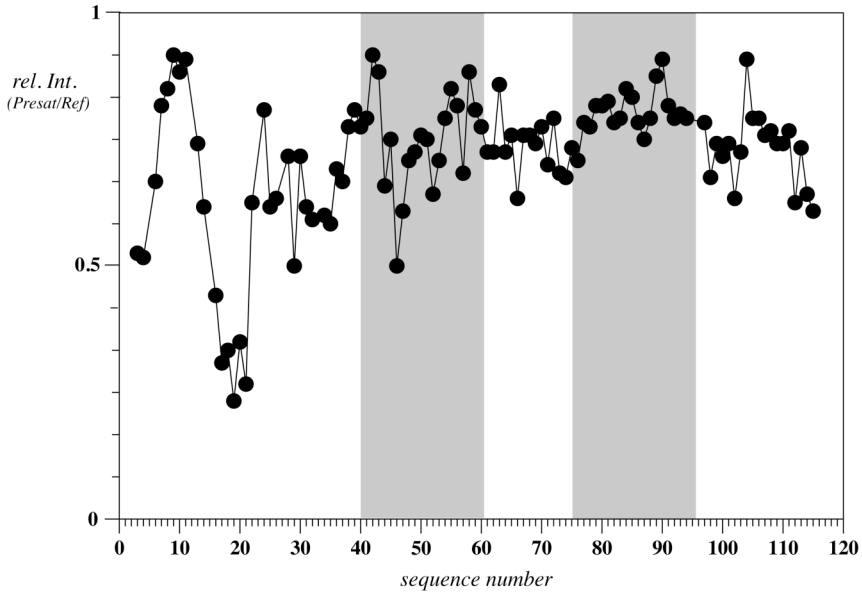


Table S3 Chemical shift of N-TM1-TM2-Y4

No.	Residue	H	N	CA	CB	CO
3	THR	8.502	116.704	64.639	68.049	175.956
4	SER	8.297	117.174	61.182	62.11	
6	LEU	7.541	118.651	57.169	40.818	178.007
7	LEU	7.824	116.372	57.203	40.259	178.092
8	ALA	7.431	118.379	53.659	17.352	179.231
9	LEU	7.319	116.245	56.134	41.323	177.463
10	LEU	7.352	115.795	55.042	42.384	175.882
11	LEU	7.336	116.954	52.441	40.46	174.681
13	LYS	8.211	120.489	55.169	32.156	176.35
14	SER	8.163	117.305	55.913	63.147	173.046
16	GLN	8.235	119.338	55.489	28.446	176.399
17	GLY	8.14	109.486	44.86		174.064
18	GLU	8.118	120.26	56.033	29.231	176.105
19	ASN	8.292	118.881	52.891	38.159	174.891
20	ARG	8.087	121.063	55.578	29.724	175.979
21	SER	8.158	116.689	57.852	63.693	173.757
22	LYS	7.993	123.056	53.762	31.71	174.26
24	LEU	8.099	120.823	54.763	41.43	177.019
25	GLY	8.042	108.218	44.654		173.293
26	THR	7.822	114.31	59.277	68.975	173.156
28	TYR	7.73	118.795	57.613	38.16	174.899
29	ASN	8.086	119.892	52.539	37.743	175.384
30	PHE	8.113	121.307	59.609	38.722	176.702
31	SER	8.225	114.721	60.226	62.833	175.609
32	GLU	7.85	120.867	57.44	28.538	177.125
33	HIS	7.831	115.255	55.926	28.369	175.614
34	CYS	7.956	118.697	60.705	27.006	175.257
35	GLN	8.248	120.432	57.827	27.72	176.773
36	ASP	8.113	118.158	55.651	39.767	177.393
37	SER	7.797	114.924	61.304	63.068	175.333
38	VAL	7.798	121.908	65.628	30.682	176.829
39	ASP	8.08	119.262	56.611	39.228	178.94
40	VAL	7.717	119.968	65.61	30.696	177.133
41	MET	7.834	118.844	58.492	31.117	177.828
42	VAL	8.409	118.005	66.193	30.847	178.527
43	PHE	7.972	121.986	60.701	37.947	177.557
44	ILE	8.253	119.943	64.547	36.878	177.638
45	VAL	8.1	114.911	64.891	30.67	176.856
46	THR	7.736	107.996	63.125	69.177	175.815
47	SER	7.486	116.533	60.129	63.251	173.66
48	TYR	7.513	119.947	57.3	38.891	174.547
49	SER	7.996	114.635	57.196	64.313	175.532
50	ILE	8.668	123.423	63.026	36.645	176.94
51	GLU	8.527	119.58	59.329	27.728	179.185
52	THR	7.736	116.742	65.512	68.076	176.22
53	VAL	7.675	120.84	66.226	30.679	177.119
54	VAL	8.366	117.993	66.304	30.539	178.266

55	GLY	7.853	107.714	46.592		176.326
56	VAL	7.824	121.858	65.569	30.632	178.171
57	LEU	8.181	119.129	57.552	40.384	179.287
58	GLY	8.615	106.718	47.107		175.71
59	ASN	7.753	120.101	55.856	37.98	177.852
60	LEU	8.223	120.884	57.618	40.632	178.803
61	CYS	8.253	117.607	63.741	26.203	176.421
62	LEU	7.779	118.724	57.244	40.254	179.732
63	MET	8.021	119.473	58.33	31.962	177.828
64	CYS	7.88	116.525	62.913	26.804	175.883
65	VAL	7.788	116.707	65.039	30.796	177.477
66	THR	7.812	114.599	65.039	68.594	176.237
67	VAL	7.85	120.146	64.688	30.548	176.854
68	ARG	7.831	119.629	57.561	28.783	177.026
69	GLN	7.846	117.188	57.086	27.852	177.34
70	LYS	7.903	119.125	57.185	31.888	177.399
71	GLU	8.139	117.887	56.43	28.092	176.762
72	LYS	7.951	120.119	57.557	31.378	176.943
73	ALA	8.009	122.21	52.929	17.831	177.742
74	ASN	8.017	116.734	53.932	38.031	176.59
75	VAL	8.101	119.059	64.495	30.794	176.585
76	THR	7.937	114.408	65.352	67.953	175.438
77	ASN	7.933	118.552	55.01	37.823	176.631
78	LEU	7.664	120.124	56.61	41.235	177.914
79	LEU	7.748	118.737	56.825	40.967	177.916
80	ILE	7.748	115.977	62.554	37.006	176.982
81	ALA	7.726	121.253	53.177	17.982	178.009
82	ASN	7.648	114.432	53.546	39.12	174.909
83	LEU	7.713	121.071	55.624	41.242	177.014
84	ALA	8.026	122.608	52.829	17.628	178.085
85	PHE	7.832	116.965	58.715	38.221	176.416
86	SER	8.013	114.77	60.676	62.804	175.522
87	ASP	8.107	121.944	56.714	39.495	178.197
88	PHE	7.93	120.164	60.276	38.171	176.854
89	LEU	7.876	118.322	57.193	40.554	178.23
90	MET	7.946	115.666	57.389	30.773	178.358
91	CYS	7.678	117.566	61.858	26.084	176.489
92	LEU	7.523	119.597	57.08	40.879	176.955
93	LEU	7.541	113.374	55.777	41.017	177.303
94	CYS	7.463	113.246	60.09	27.666	175.267
97	LEU	8.131	117.315	56.787	39.982	178.197
98	THR	7.656	112.816	65.312	68.215	
99	ALA	7.805	124.299	54.592	17.593	178.848
100	VAL	7.664	116.162	65.607	30.546	177.032
101	TYR	7.939	118.513	59.883	36.748	177.758
102	THR	7.909	114.942	65.878	68.552	176.027
103	ILE	7.801	120.128	64.101	36.779	177.296
104	MET	8.011	117.727	58.19	31.58	177.495
105	ASP	7.978	118.206	56.131	39.54	177.742
106	TYR	7.862	118.423	59.975	37.529	177.029

107	TRP	8.19	120.39	59.408	28.847	176.828
108	ILE	7.786	117.441	62.928	36.541	177.551
109	PHE	7.666	119.482	59.212	38.067	177.289
110	GLY	8.122	107.674	46.379		174.728
111	GLU	8.246	120.011	57.799	27.902	177.57
112	THR	7.774	113.838	64.716	68.348	175.617
113	LEU	7.848	121.518	56.756	40.886	178.085
114	CYS	7.807	115.664	60.692	26.926	175.433
115	LYS	7.8	119.058	57.202	31.471	176.86
116	HIS	7.882	116.16	55.42	28.445	174.38
117	HIS	7.943	117.91			173.812

Chapter 5

Recognition of neurohormones of NPY family by their receptors

5.1 Introduction

G protein-coupled receptors (GPCRs) play a pivotal role in biology by transmitting signals from outside the cell into the cell, thereby bypassing the cellular membrane that is impermeable for most of the signaling molecules. So far only a single high-resolution structure of an intact GPCR, namely from bovine rhodopsin, has been published[1]. Therein, the polypeptide chain crosses the membrane seven times, and these segments are comprised of α-helices, which pack against each other. The transmembrane-helices are connected via three extracellular and three intracellular loops. In addition, all GPCRs possess N- and C-terminal domains of varying size and may in addition be glycosylated or further modified.

Whereas at least some information on the structure of GPCRs is available, and modeling using structural information derived from rhodopsin is frequently done[2], little is known on the exact binding mode of ligands to the GPCRs. To some extent such information is available from receptor mutagenesis studies (e.g. for the case of rhodopsin see references in Rader et al.[3]), from photoaffinity labeling[4; 5] or spin labelling[6] experiments. Interestingly, the results from these techniques are not always fully consistent, which may be due to the fact that mutants with reduced affinity potentially suffer from significant distortions in their architecture even if the mutated residues are not directly involved in binding. Nevertheless, all these techniques have dramatically improved our understanding of ligand binding to GPCRs.

An interesting question related to this subject is whether the ligands to GPCRs are directly recognized from solution or whether they associate with the membrane prior to receptor binding (see Fig. 1). Kaiser and Kezdy in their seminal paper have noticed that many features of the ligands seem to be optimized for membrane binding[7]. Moreover, they realized that the ligands are often far too large for the few contacts they probably make with their receptors[8]. Accordingly, they suspected that the

hormones were evolutionarily optimized for membrane binding. This view was then further developed by Schwyzer, who established the membrane compartment model[9; 10; 11]. His theory states that the membrane also helps to sort the ligands into the correct compartments, e.g. the aqueous phase, the water-membrane interface or the membrane interior, and that it could in addition pre-organize the molecules to adopt a conformation similar to the bioactive (receptor-bound) one.

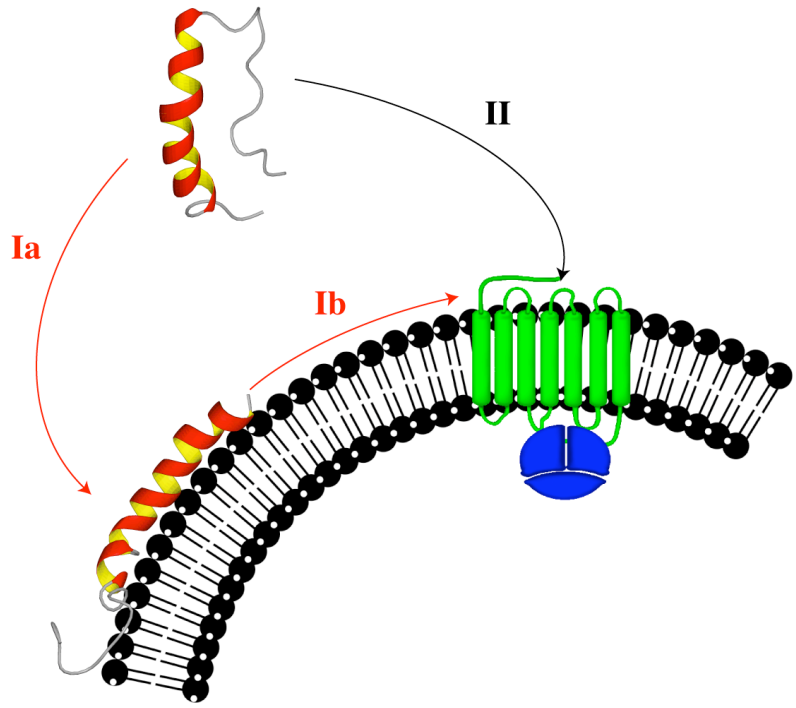


Fig.1: Pathway for recognition of hormones from the NPY family by their receptors via direct recognition from solution (path II) or via a series of events, in which association with the membrane (Ia) followed by lateral diffusion along the membrane precedes receptor binding (Ib).

The ideas of Kaiser, Kezdy and Schwyzer have been heavily debated during the last 20 years or so. Moroder in an elegant approach has added lipid chains to the peptides in order to increase their affinity for the membrane[12]. The membrane compartment model was originally formulated very generally in such that it may be applicable to all kinds of ligands to GPCRs, but in the meantime many cases have been reported in which such a pathway is unlikely. We have decided to take a structural approach to this question and look at non-modified hormones[13]. In our studies we have structurally characterized peptides from the neuropeptide Y family both in solution and in the membrane-bound states[13; 14; 15; 16; 17; 18]. We have subsequently investigated whether structural differences of a series of peptides with

known pharmacology can be better related to the solution structures or to the structures of the membrane-bound state. This view requires that molecules with highly similar pharmacology at different receptor subtypes should possess similar conformations in the state from which they are recognized. If this would not be the case the associated changes of entropic terms would result in differences in the free energies of binding, which in turn would translate into differences in binding affinities or signal transduction efficiencies.

5.2 Structural Features of Peptides from the Neuropeptide Y Family in Solution

The neuropeptide Y family comprises three different members: The neuropeptide Y (NPY), the peptide YY (PYY) and the pancreatic peptide (PP) [19]. These C-terminally amidated peptides contain 36 amino acids. Moreover, the N-terminal segment contains a Pro-rich segment, and in addition many aromatic residues, in particular Tyr residues, are found in the sequence, which is also the reason for the name (NP**Y**). The peptides target a heterologous population of G protein-coupled receptors, the so-called Y receptors[20; 21]. These are approx. 375 to 450 amino acid long polypeptides coupled to proteins of the G_i subtype. So far, four Y receptor subtypes have been identified, sequenced and pharmacologically characterized. They occur in the central or peripheral nervous system and in the gastrointestinal tract. Activation of these receptors by one of the peptides triggers pharmacologically highly important functions[22] such as vasoconstriction, memory retention and food uptake, which recently moved into focus of research[23; 24].

The pancreatic polypeptide was one of the first smaller peptides that is devoid of disulfide bonds or any other rigidifying modifications, for which a crystal structure became available. In their work Blundell *et al.* could demonstrate that the C-terminal half of the molecule, comprising residues 14-29, forms an amphiphatic α-helix, which via a turn encompassing residues 10 to 13 is connected to an N-terminal type II polyproline helix[25]. The latter forms a hydrophobic contact by back-folding onto the amphiphatic α-helix resulting in a structural motif that became known as the PP fold. The crystal structure of avian PP (aPP) also revealed the presence of a dimer. Therein, aromatic residues form a hydrophobic core, in which the π-systems of Tyr and Phe residues stack onto each other thereby mutually stabilizing both the back-fold as well as the dimeric nature of the peptides. Later on, it could be demonstrated by NMR that in case of bovine PP (bPP) the PP-fold also exists in solution[26]. The neuropeptide Y was originally also proposed to display a PP-fold type structure[27]. However, a later, better resolved structure revealed that the N terminus is flexible[28], and this view was then supported by our data on backbone dynamics[18] and from spin-labeling studies[29]. PYY is again back-folded and in its structure highly similar to PP[16; 30] (see Fig. 2), although PYY is believed to largely exists in monomeric form[31]. Common to all

molecules of the NPY family is the C-terminal α-helix, but differences exist in the structure of the C-terminal pentapeptide.

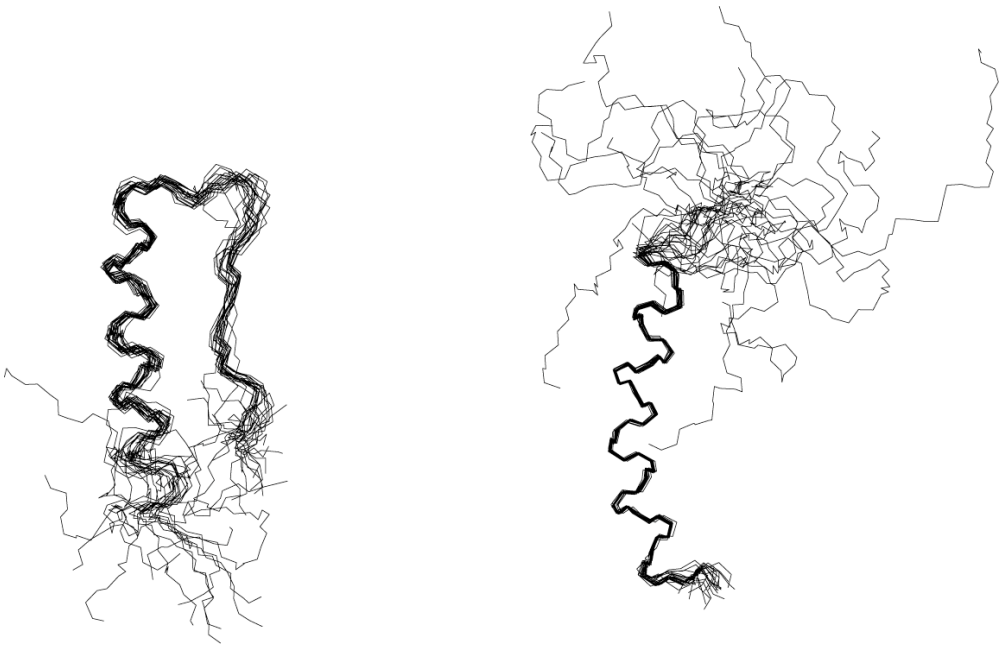


Fig.2: Backbone representation of the 20 lowest-energy conformers of PYY as determined by NMR spectroscopy in solution (left) or when bound to DPC micelles (right).

Most of the peptides from the NPY family are dimeric in nature with dissociation constants in the low μM range for NPY and PP, with the exact values very much depending on the pH. PYY in contrast is proposed to largely exist in monomeric form. We have introduced a particular amino acid derived from the spin label TEMPO in combination with ^{15}N labeling to establish spatial proximity of residues of NPY in the dimer[18]. From these data we concluded that in the NPY dimer both parallel and anti-parallel association of the helices occurs. However, it is presently unclear whether contacts are present simultaneously by forming a three-helix bundle, or whether the two association modes interconvert via a partially unfolded state.

It is somewhat surprising that NPY and PYY differ so much in their solution structure considering that their sequence homology is larger than 80%. In particular, the distribution of hydrophobic, polar or charged residues along the sequence is almost identical. In an attempt to better understand the molecular features that favor one conformation over the other we have expressed all single Pro-to-Ala mutants as well as certain mutants, in which Tyr was replaced by Ala. One of the most obvious differences between sequences of NPY and PYY is that Pro in NPY occurs at position

13 whereas it is found at position 14 in PYY. Replacing Pro-14 in PYY by Ala did not disrupt the back-fold. In contrast, when Pro-14 was shifted to position 13, where it is found in NPY, the back-fold is no longer present, and hence a single amino acid shift by one position seems to be sufficient to transform the molecule from the PP-fold conformation into the non back-folded species. Replacement of Pro residues 2, 5 and 8, which are part of the polyproline type II helix, by Ala always destabilized the back-fold dramatically (unpublished results), and we believe that this is due to an increase of the entropy of the non back-folded species. In addition, Tyr7 in bPP is involved in the π-stacking interactions, but no such Tyr residue is found in position 7 in PYY, although the structure of PYY is highly similar to the one of bPP. We have introduced Tyr into position 7 in PYY, and noticed formation of a stable dimer, in which ring-stacking interactions almost identical to the ones observed in PP occur.

5.3 Structural Features of Peptides from the Neuropeptide Y Family in the Membrane-bound State

Biophysical properties of the membrane interior or the water-membrane interface are substantially different from those in bulk aqueous solution, and partitioning of peptides into these compartments may result in structural changes. Accordingly, we have determined structures of the NPY peptides in the presence of membrane-mimicking dodecylphosphocholine (DPC) micelles[32]. These readily assemble into spherical aggregates, the size of which is still compatible with high-resolution NMR studies.

Fig. 2 displays a comparison of the structure of procine PYY (pPYY) in aqueous solution and in the presence of DPC micelles. Clearly, a remarkable conformational transition has occurred: Whereas the C-terminal α-helix has been retained, the contact between the N- and C-terminal segments has been disrupted[16]. The hydrophobic side of the α-helix now forms the membrane-binding interface. The membrane also serves to stabilize the C-terminal α-helix in NPY and in particular the C-terminal pentapeptide, which is rather flexible in solution but very rigid in the micelle-bound state[18]. We have mainly used a micelle-integrating spin-label, 5-doxylstearate, which in a distance-dependent manner broadens signals from protons close to the phospholipid head-groups, in order to establish the membrane-binding topology. All peptides from the NPY family are anchored onto the micelle through insertion of side-chains from hydrophobic residues into the hydrophobic interior. More polar or charged residues are oriented towards the aqueous phase. The N-terminal segment, which is rather rigid in the back-folded peptides, always becomes flexible in the micelle-bound state and mainly diffuses freely in solution.

The structural transition can be well explained using thermodynamic arguments: Wimley and White have determined free energies of transferring whole amino acids from bulk solution into the membrane-water interface or into the membrane interior [33]. Their data confirm our intuitive understanding of how polarity and hydrophobicity should influence partitioning. In particular, their work, which has been recently verified based on a biological read-out of partitioning[34], attributes a particular role to aromatic amino acids, and in particular to Tyr and Trp residues[35]. Predicting the membrane-anchoring mode of the NPY peptides based on their thermodynamic data accurately reproduces our experimental findings[13]. For example, the N-terminus of

NPY or PYY makes no contacts with the micelle surface, whereas such contacts are observed in the case of bPP, an observation that can be readily attributed to the presence of Tyr7 in bPP. The energies of transferring whole amino acids into the water-membrane interface are also systematically lower in the C-terminal part of the molecules (the α-helix) compared to the N-terminal part, and therefore one would predict that it is the hydrophobic side of the C-terminal helix that binds to the micelles. We like to emphasize that this supports the view that these molecules have been evolutionarily optimized for membrane binding.

I like to summarize our structural studies of peptides from the NPY family in the micelle-bound state as follows:

i) all peptides that adopt the PP fold in solution do not display the association of the N- and C-terminal segments in the micelle-bound state. This hydrophobic contact is replaced by binding of the helix to the membrane.

ii) the peptides are oriented such that the hydrophobic side of the amphiphatic helix constitutes the membrane-binding interface. Side-chains of hydrophobic residues penetrate into the membrane interior. A particular role is found for the Tyr residues, which always partition into the water-membrane interface and probably contribute to overall membrane-binding substantially. The N terminus is generally unstructured but associates with the micelle surface when it contains aromatic residues.

iii) the C-terminal pentapeptide containing Arg33 and Arg35, for which a direct contact with the Y receptors has been postulated, becomes structurally better defined upon micelle binding. Peptides, for which large changes in pharmacology with respect to NPY are observed, often display a significantly different conformation in that region.

iv) we find no indication that the peptides, which mostly exist in dimeric form in solution, still form aggregates in the micelle-bound state.

NPY and PYY have very high sequence similarity. Moreover, their pharmacology is almost identical: Both peptides are ligands with nanomolar binding affinities at all receptor subtypes. Yet, the structure of the two peptides in solution is very different [16]: In NPY the N terminus is unstructured, whereas it is back-folded in PYY. We expect that peptides with very similar pharmacology *at all receptor subtypes* should also be conformationally similar, because the are likely to share a common binding mode. In the micelle-bound state the structural difference that exists between NPY

and PYY in solution is removed because the back-fold is disrupted, and the conformations of the two peptides in this environment become virtually indistinguishable. In addition, bPP, a peptide highly selective to the Y_4 receptor subtype, which in solution shares much more similarity with PYY than NPY with PYY, differs significantly in the C-terminal pentapeptide in the micelle-bound state. A comparison of the conformations of PYY, NPY and PP in the two environments is depicted in Fig. 3. The differences in structure of NPY/PYY with PP are accompanied by differences in the mode by which these peptides bind to the micelles: In case of NPY and PYY no association of the N terminus with the micelle is observed, whereas binding of the N-terminal segment of PP (basically via partitioning of Tyr-7 into the interface) to the micelle is detected. To conclude, similarities and differences of pharmacology of NPY, PYY and PP are much better correlated in the micelle-bound state.

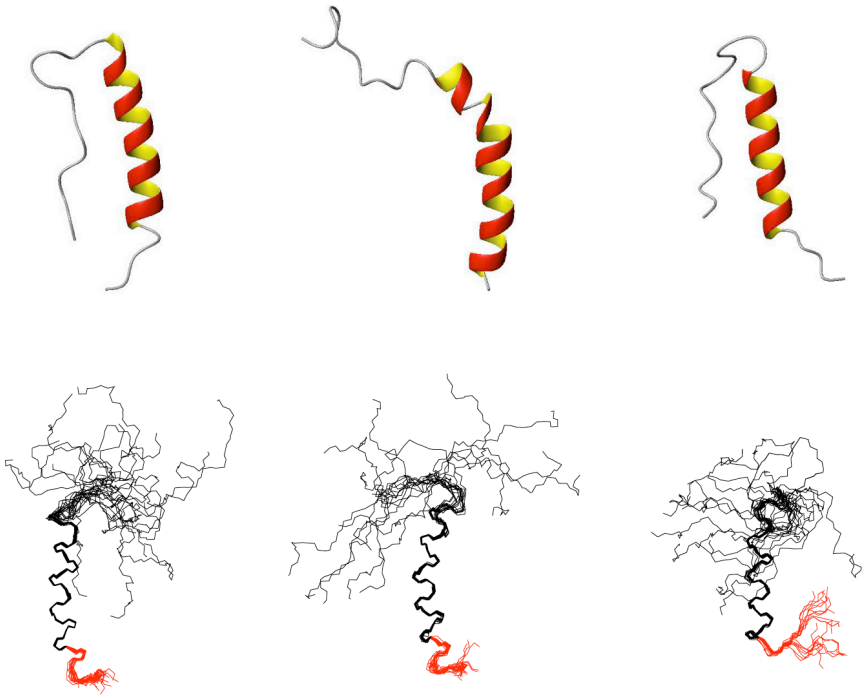


Fig.3: Comparison of the structures ob pPYY (left), pNPY (middle) and bPP (right) in solution (top panel) and in the micelle-bound state (bottom panel). The C-terminal pentapeptide is color-coded in red. Figure reproduced from Lerch et. al.[16].

5.4 A Series of Events for Receptor Binding

The experimental results described in the previous section indicate that hormones from the neuropeptide Y family are not directly recognized from solution, but rather from the membrane-bound state. Unfortunately, no structural data of the complex formed between the peptides and any of the Y receptor subtypes are presently available. Some mutagenesis data have been published for hNPY at the human Y1 receptor. Walker *et al*. have postulated a prominent role for Asp residues for binding[36]. In particular, an Asp residue at the interface between the 6th TM domain and the third extracellular loop is conserved in all known Y receptor sequences. Moreover, interacting partners were proposed to reside in the N-terminal domain as well as in the first extracellular loop[37]. Beck-Sickinger *et al.* have conducted a full Ala scan of hNPY at the human Y1 receptor [38]. Most significant reductions in affinity were observed for Arg33 and Arg35. Moreover, binding was almost completely abolished when the C-terminal amide was replaced with a free C terminus. From these results direct interactions involving hormone residues Arg33 and/or Arg35 and an Asp receptor residue (Asp289 in case of the hY1 receptor) have been postulated. Nevertheless, Dougherty has also proposed that π-cation interactions involving one of the Arg residues and aromatic receptor residues contribute to binding[39]. Even if the exact nature of the ligand-receptor complex is presently unknown, partitioning of the hormones into binding pockets located deeper in the membrane is unlikely considering the presence of a larger number of polar or charged residues in the hormones and the absence of corresponding polar residues in the membrane interior of the Y receptors. Hence, it is most likely that NPY is binding to the extracellular loops and/or to the N-terminal domain[40].

We are proposing the following series of events during binding of hormones from the NPY family to the Y receptors [13; 16] as described in Fig. 4: i) initially, association of the peptides with the membrane occurs. This event is driven by electrostatic attraction involving cationic residues of the peptides and the negative charges of the phospholipids. Such electrostatic interactions have also been observed in case of the membrane-active peptides, e.g. cell-penetrating or antimicrobial peptides, and were characterized in more detail using surface plasmon resonance (SPR)[41] or calorimetric techniques. Once the peptides have approached the membrane-surface closely enough they reorient such that the side-chains of the hydrophobic residues penetrate into the

membrane interior. The peptides then diffuse laterally along the membrane surface towards the receptors, and there may also be an electrostatic component involved in the latter. It is important in that respect that the biophysics in such an environment are different from those encountered in bulk solution[42]: There is a steep gradient in charge density along the membrane normal. Depending on its exact location the dielectric constant may vary considerably, from about 78 in bulk solution to about 2 in the membrane center, and hence solvation of polar groups could be very different. A consequence of this is the often-observed partitioning-folding by which peptides that are unstructured in solution adopt a unique conformation when partitioning into the interface[42]. The lowered dielectric constant also results in an increase of the range for which electrostatic interactions are relevant due to the diminished Debye-Hückel screening. It is therefore reasonable to assume that electrostatics will be more important in an environment of reduced permittivity. Again, considering the steep gradient of charges along the membrane normal, the importance of the mentioned arguments depends very much on where the peptide is exactly located in the membrane (interface).

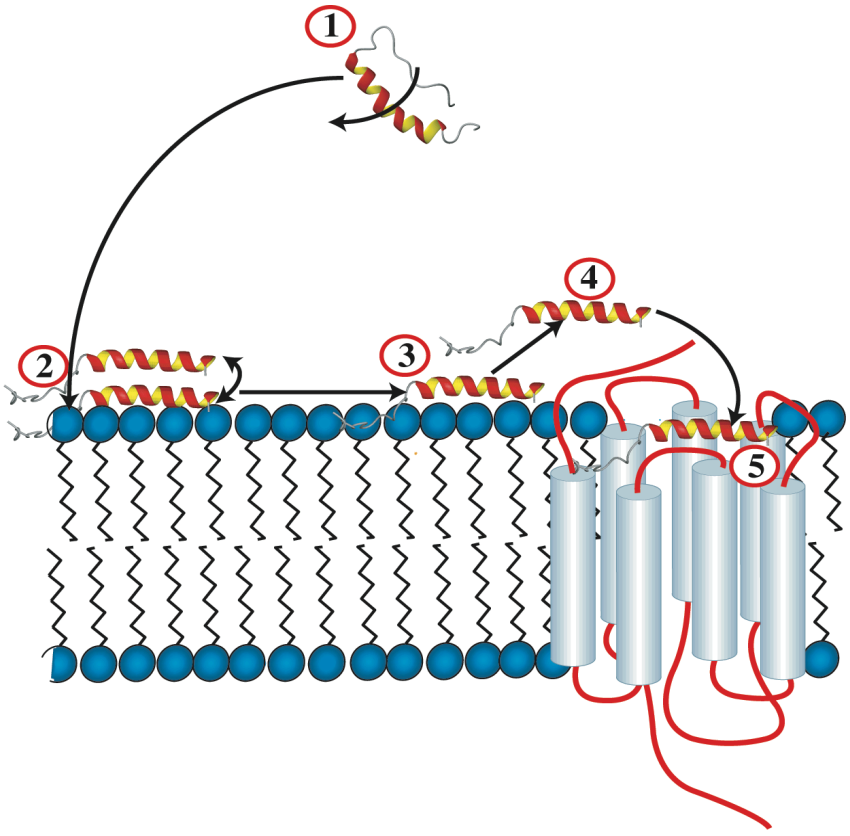


Fig.4: Detailed mechanism of receptor binding including membrane association (1), lateral diffusion towards the receptor (2), dissociation from the membrane (3) possibly linked to (transient) binding to the N-terminal domain of the receptor (4) followed by diffusion into the binding pocket (5).

Another aspect of receptor binding is concerned with how these peptides can possibly diffuse into the binding pocket from the membrane-associated state. It is unlikely that they can directly enter the binding site, because the receptor-surrounding lipids and the loop regions are separated in space and most likely do not mutually penetrate. Moreover, at least in the case of rhodopsin the scaffold built by the 7 TM bundle is fairly rigid[1], and any pathway that requires the scaffold to re-organize in order to allow entry of the ligands is unlikely. It is therefore conceivable that the hormones must come off the membrane in order to diffuse into the binding pockets. In order to be able to do so, the affinity for binding to the membrane should not be too high. We have determined affinities of NPY, PYY and PP to neutral and negatively charged phospholipids surfaces using surface-plasmon resonance (SPR, BiaCore)[14]. The measured association constants to zwitterionic phospholipids were in the range between $4.6 \cdot 10^4$ for PP and $6.8 \cdot 10^4$ for NPY, and hence the peptides will be in dynamic equilibrium between the membrane-bound form and a dissociated state, which, however, remains close to the membrane surface. These values are significantly lower than those observed for membrane-active peptides such as melittin.

5.5 Investigating the Structural Transition between Bulk Solution and the Membrane Environment

Since we have emphasized the importance of the membrane-bound state for receptor recognition it is of much interest to know whether the peptides, once they come off the membrane, immediately adopt the conformation observed in bulk solution, or whether the conformation of the membrane-bound species is retained to some extent. We have recently realized that the conformation of PYY in methanol almost perfectly superimposes with the structure of the micelle-bound state. Fig. 5 displays a comparison of the structures of PYY in methanol and in its DPC-micelle bound state. We have recently made extensive use of the heteronuclear NOE to characterize backbone dynamics and to quantify the extent of back-folding. Both, structural data and backbone dynamics indicate that water-methanol mixtures may be very useful to simulate the transition occurring in the solvent environment when the peptides diffuse from the aqueous phase to the water-membrane interface. We therefore recorded spectra of a variant of PYY, Y7-PYY, in various water-methanol mixtures. As a result we noticed a smooth transition between the two forms, e.g. the state in solution and in the micelle-bound form. Moreover, we have seen that the population of the non back-folded form is still high, even when the water content is as high as 60%. From these data we conclude that the conformation of the membrane-bound species is retained as long as the peptides don't diffuse back into bulk solution but remain in vicinity of the interface where solvent properties, in particular the dielectric constant, have not been changed too dramatically. We believe that the binding constants for the membrane are sufficiently high, so that re-binding occurs quickly preventing the peptides from diffusing back into bulk solution.

Following the arguments presented above we conclude that the conformation of the membrane-bound hormone is important, even when the peptides need to detach from the membrane in order to diffuse into the binding pocket.

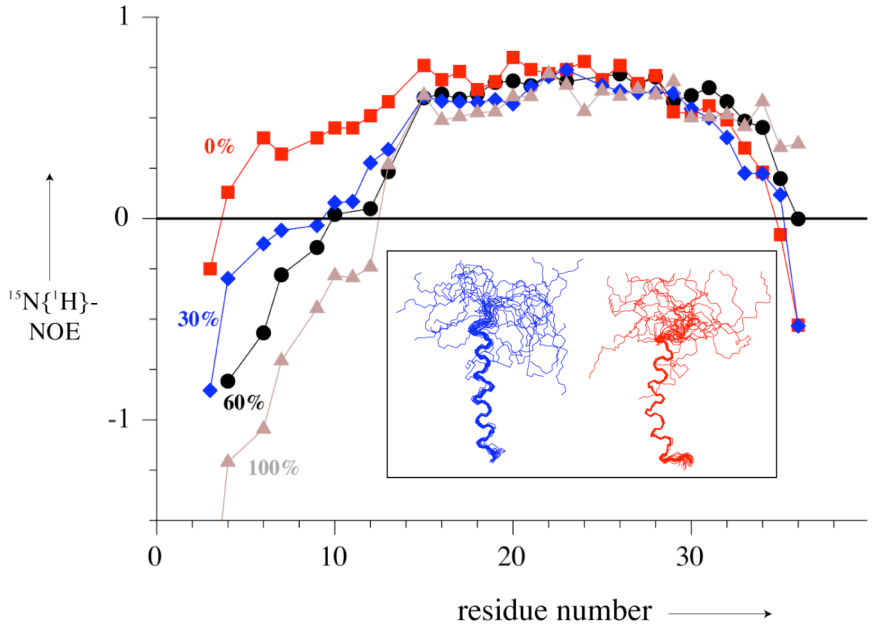


Fig.5: Values of the ^{15}N{^{1}H}-NOE of pPYY in various water-methanol mixtures. The structures of pPYY in methanol and in the micelle-bound state are depicted as small insets in blue and red, respectively.

5.6 Removal of the Back-fold When Diffusing towards the Membrane Occurs Cooperatively

In the previous section the structural transition occurring in PYY when it diffuses from bulk solution into the water-membrane interface was described and data were presented to demonstrate that water-methanol mixtures are useful to simulate that event. One remaining interesting question is whether unfolding occurs in a non-concerted manner like unzipping, e.g. by removing individual contacts of proline residues with the C-terminal helix from one end onwards, or whether all these contacts are disrupted simultaneously. Replacement of Pro residues at positions 2, 5 and 8 has resulted in removal of the PP fold, and hence it is likely that each of these residues contributes to the back-fold. However, each of these interactions on its own is too weak to stabilize the PP fold. If this is the case then only cooperative folding can result in a stable structure, because the latter would require all contacts to be formed simultaneously.

Fig. 6 displays the chemical shifts changes of selected amide protons, and the data have been grouped into residues being part of the interface formed by the PP-fold, and those residing on the side of the helix that is opposite to the interface. Clearly, chemical shifts of amide protons located at the interface display a sigmoidal shape, and the point of inflection occurs at about 40% methanol. Given that the shapes of the curves from all amide protons located at the interface are similar, and considering the fact that the point of inflection is roughly the same in all cases, these data clearly confirm the view that formation of the PP fold is cooperative. Amide protons pointing to the opposite side will only experience a change in solvent properties. In this case, increasing methanol concentrations leads to a stabilization of the internal hydrogen bonds resulting in low-field shifts of these protons.

Moreover, we have noticed that changes in backbone dynamics do not seem to be well correlated to changes in chemical shift in these solvent mixtures (data not shown). While this is an issue of present investigations, it indicates that a smoother transition between a fully flexible N terminus and one that is tightly fixed by back-folding, occurs. It is of particular interest that these experiments can be performed using the same molecule and hence do not require the preparation of a series of mutant peptides.

Moreover, in comparison to methods in which changes in fluorescence are observed, it does not require introducing aromatic residues into specific (non-native) positions, which themselves will influence membrane partitioning and, possibly also to some extent, folding.

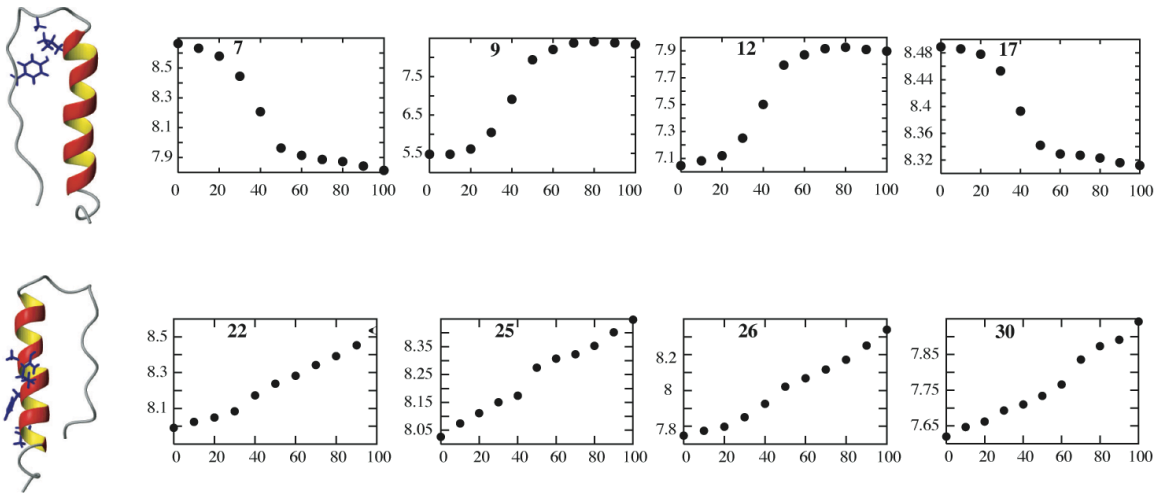


Fig.6: Chemical shifts of amide protons of selected residues of Tyr7-PYY in dependence of the methanol content in water-methanol mixtures depicted for residues part of the back-fold interface (top row) and for residues placed at other positions (bottom row). Side-chains of these residues have been drawn in the structures depicted on the left.

5.7 Weak Interactions between the Hormones and the N-terminal Domains Are Observed

It was already known from mutagenesis experiments that residues of the N terminal domains of the Y receptors might be involved in binding[40]. All of the Y receptor subtypes contain such domains which comprise between 42 and 53 amino acids, and which are placed on the extracellular side. It is therefore conceivable that the hormones may actually in their initial encounter with the receptors, at least transiently, bind to these domains. Such an event may also help to shift the equilibrium between membrane-associated hormomes and those that have come off the membrane towards the latter. In that way, the N-terminal domains may help to transfer the peptides from the micelle-bound state into the binding pockets. We have synthesized or expressed the N-terminal domains from all Y receptor subtypes and tested binding of NPY, PYY and PP to them. In all cases we could detect such an interaction by chemical shift mapping. The changes in peak positions in the $^{15}N,^{1}H$-correlation experiment for all peptides when interacting with the human Y2 receptor are depicted in Fig. 7. Clearly, specific changes are observed for resonances of the hormones. In contrast, there are many more changes detected for the Y2 receptor fragment. Dynamics data indicate that the isolated Y2 N-terminal fragment is largely unfolded, and we believe that the changes in chemical shift observed for resonances of the Y2 fragment more likely indicate that it becomes at least partially structured upon interaction with the hormones. The profiles of a particular peptide, as demonstrated for bPP in Fig. 7, are also different at the various receptor subtypes. We are currently in the process of characterizing these interactions in much more detail, both structurally as well as with respect to binding affinities or residues involved in forming the contacts.

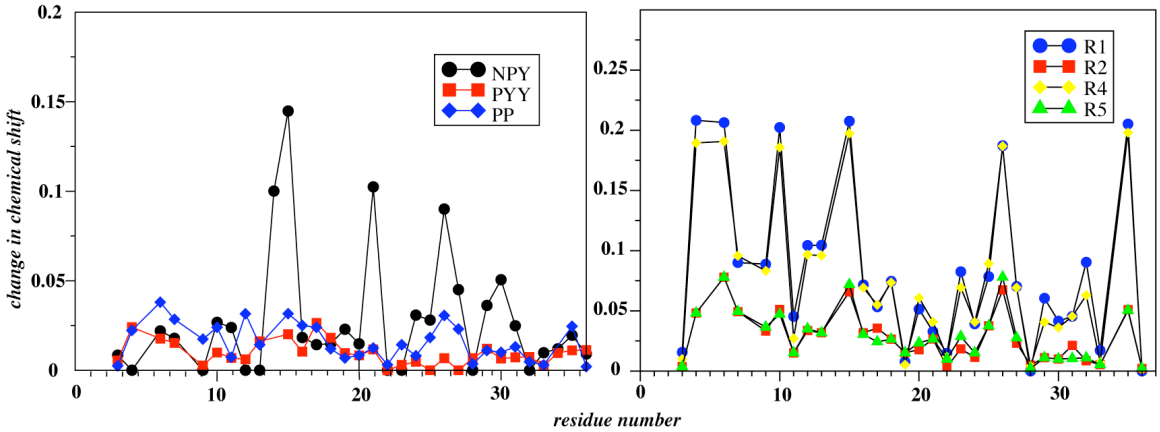


Fig.7: Results from chemical shift mapping of pNPY, pPYY and bPP at the N-terminal domain of the Y2 receptor (left) or for bPP at all Y receptor subtypes (right).

5.8 Outlook

Structural biology of receptor-ligand interactions is still a field in which much more information is desperately needed. In particular, larger quantities of purified and successfully reconstituted receptors are required for structural studies, and much progress in that area is presently made [43; 44]. Crystallization of membrane proteins for use in crystallography is still difficult but again the numbers of published structures is rapidly expanding. NMR methodology has witnessed a number of recent improvements such as TROSY techniques[45] or the use of residual dipolar couplings[46] that have dramatically increased the maximum molecular weight at which structures of such molecules may still be elucidated. In addition, rapid progress is made in the solid-state NMR field[47]. It is therefore conceivable that structures of recombinantly expressed GPCRs will be solved in near future by one of these techniques, provided that enough (labeled) protein can be expressed.

However, even when such structures (or even those of GPCR-ligand complexes) become available the pathway of binding may still not be easily deduced. We have previously proposed that all events during receptor recognition are important, and that highly effective ligands need to be optimized for all steps occurring during recognition[13]. I like to emphasize here that this does not imply that all ligands follow the same path, e.g. the small molecule drugs certainly do not contain all the functionality for such a pathway. Nevertheless, it is of much importance for our understanding of the basic biology behind receptor function how these naturally occurring systems operate. It has been known for some time that biological membranes are chemically inhomogeneous entities. Much attention has been recently paid to lipid rafts[48], and it has been proposed that ligands binding to GPCRs may actually be influenced by the presence of the latter. Moreover, the morphology of the membrane may depend on the state of the cell. A pathway that includes membrane association may therefore be influenced by changes in membrane morphology, and the question whether hormones associate with the membrane prior to receptor binding could therefore have more implications than those that are immediately obvious. Clearly, there is still a need for clever cell-biological experiments to investigate this question from a different point of view in the future.

5.9 Acknowledgements

This review contains much data that are part of the PhD thesis of Dr. Reto Bader and Dr. Mirjam Lerch, and their large contribution for the development of the described ideas was very important. We like to further thank Karin Weber, Verena Gaffner, Garielle Meininghaus-Rytz, Barbara Christen and Margot Mayrhofer for data resulting from their diploma thesis. We appreciate valuable discussion with Annette G. Beck-Sickinger, Gerd Folkers and John A. Robinson on the subject and access to their laboratory infrastructure. This project was financially supported from the Swiss National Science Foundation (SNF grant no. 3100A0-100462), the Forschungskredit der Universität Zürich (grant no. 57132601) and the Forschungskredit der ETH Zürich (grant no. 0 20 439-97 and TH-39/00-3).

5.10 References

1. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739-45.
2. Flower, D. R. (1999). Modelling G-protein-coupled receptors for drug design. *Biochim. Biophys. Acta* **1422**, 207-34.
3. Rader, A. J., Anderson, G., Isin, B., Khorana, H. G., Bahar, I. & Klein-Seetharaman, J. (2004). Identification of core amino acids stabilizing rhodopsin. *Proc Nat Acad Sci U S A* **101**, 7246-51.
4. Henry, L. K., Khare, S., Son, C., Babu, V. V., Naider, F. & Becker, J. M. (2002). Identification of a contact region between the tridecapeptide alpha-factor mating pheromone of Saccharomyces cerevisiae and its G protein-coupled receptor by photoaffinity labeling. *Biochemistry* **41**, 6128-39.
5. Rihakova, L., Deraet, M., Auger-Messier, M., Perodin, J., Boucard, A. A., Guillemette, G., Leduc, R., Lavigne, P. & Escher, E. (2002). Methionine proximity assay, a novel method for exploring peptide ligand-receptor interaction. *J Recept Signal Transduct Res* **22**, 297-313.
6. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. (1990). Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science* **248**, 1088-92.
7. Kaiser, E. T. & Kezdy, F. J. (1983). Secondary structures of proteins and peptides in amphiphilic environments. *Proc Natl Acad Sci U S A* **80**, 1137-43.
8. Kaiser, E. T. & Kezdy, F. J. (1984). Amphiphilic secondary structure: design of peptide hormones. *Science* **223**, 249-55.
9. Sargent, D. F. & Schwyzer, R. (1986). Membrane lipid phase as catalyst for peptide-receptor interactions. *Proc Natl Acad Sci U S A* **83**, 5774-8.
10. Schwyzer, R. (1986). Molecular mechanism of opioid receptor selection. *Biochemistry* **25**, 6335-42.
11. Schwyzer, R. (1995). In search of the 'bio-active conformation'--is it induced by the target cell membrane? *J Mol Recognit* **8**, 3-8.
12. Moroder, L., Romano, R., Guba, W., Mierke, D. F., Kessler, H., Delporte, C., Winand, J. & Christophe, J. (1993). New evidence for a membrane-bound pathway in hormone receptor binding. *Biochemistry* **32**, 13551-9.
13. Bader, R. & Zerbe, O. (2005). Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state? *ChemBioChem* **6**, 1520-34.
14. Lerch, M., Kamimori, H., Folkers, G., Aguilar, M.-I., Beck-Sickinger, A. G. & Zerbe, O. (2005). Strongly Altered Receptor Binding Properties in PP and NPY Chimera are Accompanied by Changes in Structure and Membrane Binding. *Biochemistry* **44**, 9255 - 9264.
15. Bader, R., Rytz, G., Lerch, M., Beck-Sickinger, A. G. & Zerbe, O. (2002). Key motif to gain selectivity at the neuropeptide Y5-receptor: structure and dynamics of micelle-bound [Ala31, Pro32]-NPY. *Biochemistry* **41**, 8031-42.
16. Lerch, M., Mayrhofer, M. & Zerbe, O. (2004). Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors. *J Mol Biol* **339**, 1153-68.

17. Lerch, M., Gafner, V., Bader, R., Christen, B., Folkers, G. & Zerbe, O. (2002). Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles. *J Mol Biol* **322**, 1117-33.
18. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. (2001). Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection. *J Mol Biol* **305**, 307-392.
19. Larhammar, D. (1996). Evolution of neuropeptide Y, peptide YY and pancreatic polypeptide. *Regul Pept* **62**, 1-11.
20. Michel, M. C., Beck-Sickinger, A. G., Cox, H., Doods, H. N., Herzog, H., Larhammar, D., Quirion, R., Schwartz, T. & Westfall, T. (1998). XVI. International Union of Pharmacology recommendations for the nomenclature of neuropeptide Y, peptide YY, and pancreatic polypeptide receptors. *Pharmacol Rev* **50**, 143-50.
21. Berglund, M. M., Hipskind, P. A. & Gehlert, D. R. (2003). Recent developments in our understanding of the physiological role of PP-fold peptide receptor subtypes. *Exp Biol Med* **228**, 217-44.
22. Turton, M. D., O'Shea, D. & Bloom, S. R. (1997). In *Neuropeptide Y and drug development* (Grundemar, L. & Bloom, S. R., eds.), pp. 15-39. Academic Press.
23. Batterham, R. L., Cowley, M. A., Small, C. J., Herzog, H., Cohen, M. A., Dakin, C. L., Wren, A. M., Brynes, A. E., Low, M. J., Ghatei, M. A., Cone, R. D. & Bloom, S. R. (2002). Gut hormone PYY3-36 physiologically inhibits food intake. *Nature* **418**, 650-654.
24. Tschop, M., Castaneda, T. R., Joost, H. G., Thone-Reineke, C., Ortmann, S., Klaus, S., Hagan, M. M., Chandler, P. C., Oswald, K. D., Benoit, S. C., Seeley, R. J., Kinzig, K. P., Moran, T. H., Beck-sickinger, A. G., Koglin, N., Rodgers, R. J., Blundell, J. E., Ishii, Y., Beattie, A. H., Holch, P., Allison, D. B., Raun, K., Madsen, K., Wulff, B. S., Stidsen, C. E., Birringer, M., Kreuzer, O. J., Schindler, M., Arndt, K., Rudolf, K., Mark, M., Deng, X. Y., Whitcomb, D. C., Halem, H., Taylor, J., Dong, J., Datta, R., Culler, M., Craney, S., Flora, D., Smiley, D. & Heiman, M. L. (2004). Physiology: does gut hormone PYY3-36 decrease food intake in rodents? *Nature* **430**, 1 p following 165; discussion 2 p following 165.
25. Blundell, T. L., Pitts, J. E., Tickle, S. P. & Wu, C. W. (1981). X-ray analysis (1.4 A resolution) of avian pancreatic polypeptide: small globular protein hormone. *Proc Natl Acad Sci U S A* **78**, 4175-9.
26. Li, X. A., Sutcliffe, M. J., Schwartz, T. W. & Dobson, C. M. (1992). Sequence-specific ^{1}H NMR assignments and solution structure of bovine pancreatic polypeptide. *Biochemistry* **31**, 1245-53.
27. Darbon, H., Bernassau, J. M., Deleuze, C., Chenu, J., Roussel, A. & Cambillau, C. (1992). Solution conformation of human neuropeptide Y by ^{1}H nuclear magnetic resonance and restrained molecular dynamics. *Eur J Biochem* **209**, 765-71.
28. Monks, S. A., Karagianis, G., Howlett, G. J. & Norton, R. S. (1996). Solution structure of human neuropeptide Y. *J Biomol NMR* **8**, 379-90.
29. Bettio, A., Gutewort, V., Poppl, A., Dinger, M. C., Zschornig, O., Klaus, A., Toniolo, C. & Beck-Sickinger, A. G. (2002). Electron paramagnetic resonance

backbone dynamics studies on spin-labelled neuropeptide Y analogues. *J Pept Sci* **8**, 671-82.
30. Keire, D. A., Kobayashi, M., Solomon, T. E. & Reeve, J. R., Jr. (2000). Solution structure of monomeric peptide YY supports the functional significance of the PP-fold. *Biochemistry* **39**, 9935-42.
31. Keire, D. A., Mannon, P., Kobayashi, M., Walsh, J. H., Solomon, T. E. & Reeve, J. R., Jr. (2000). Primary structures of PYY, (Pro^{34})PYY, and PYY-(3-36) confer different conformations and receptor selectivity. *Am J Physiol Gastrointest Liver Physiol* **279**, G129-G131.
32. Henry, G. D. & Sykes, B. D. (1994). Methods to study membrane protein structure in solution. *Meth Enzymol* **239**, 515-35.
33. Wimley, W. C. & White, S. H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Struct Biol* **3**, 842-848.
34. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. & von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-81.
35. Killian, J. A. & von Heijne, G. (2000). How proteins adapt to a membrane-water interface. *TIBS* **25**, 429-434.
36. Walker, P., Munoz, M., Martinez, R. & Peitsch, M. C. (1994). Acidic residues in extracellular loops of the human Y1 neuropeptide Y receptor are essential for ligand binding. *J Biol Chem* **269**, 2863-9.
37. Sautel, M., Rudolf, K., Wittneben, H., Herzog, H., Martinez, R., Munoz, M., Eberlein, W., Engel, W., Walker, P. & Beck-Sickinger, A. G. (1996). Neuropeptide Y and the nonpeptide antagonist BIBP 3226 share an overlapping binding site at the human Y1 receptor. *Mol Pharmacol* **50**, 285-92.
38. Beck-Sickinger, A. G., Wieland, H. A., Wittneben, H., Willim, K. D., Rudolf, K. & Jung, G. (1994). Complete L-alanine scan of neuropeptide Y reveals ligands binding to Y1 and Y2 receptors with distinguished conformations. *Eur J Biochem* **225**, 947-58.
39. Zacharias, N. & Dougherty, D. A. (2002). Cation-pi interactions in ligand recognition and catalysis. *Trends Pharmacol Sci* **23**, 281-7.
40. Beck-Sickinger, A. G. & Jung, G. (1995). Structure-activity relationships of neuropeptide Y analogues with respect to Y1 and Y2 receptors. *Biopolymers* **37**, 123-42.
41. Mozsolits, H. & Aguilar, M. I. (2002). Surface plasmon resonance spectroscopy: an emerging tool for the study of peptide-membrane interactions. *Biopolymers* **66**, 3-18.
42. Popot, J. L. & Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922.
43. Grisshammer, R., White, J. F., Trinh, L. B. & Shiloach, J. (2005). Large-scale expression and purification of a G-protein-coupled receptor for structure determination -- an overview. *J Struct Funct Genomics* **6**, 159-63.
44. Sarramegn, V., Muller, I., Milon, A. & Talmont, F. (2006). Recombinant G protein-coupled receptors from expression to renaturation: a challenge towards structure. *Cell Mol Life Sci* **63**, 1149-64.
45. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift

anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-71.
46. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111-4.
47. Opella, S. J. & Marassi, F. M. (2004). Structure determination of membrane proteins by NMR spectroscopy. *Chem Rev* **104**, 3587-606.
48. Simons, K. & Toomre, D. (2000). Lipid rafts and signal transduction. *Nat Rev Mol Cell Biol* **1**, 31-9.

Curriculum Vitae

Personal

Name:	ZOU Chao
Gender:	Male
Date of birth:	September 4, 1976
Place of birth:	Jilin Province, P.R. China
Nationality:	P.R. China
Marital status:	Married

Education

1983-1989 Jilin No.4 Primary School, Jilin Province, China.

1989-1992 Jilin No. 7 Middle School, Jilin Province, China.

1992-1995 Jilin No.1 High Middle School, Jilin Province, China.

1995-1999 Bachelor at Department of Biochemical Engineering, Beijing University of Chemical Technology, China.

2000-2003 Master at Department of Biochemical Engineering, Tianjin University, China.

Preparation and separation of phosphorylated serine peptides from casein hydrolysate.

Supervisor: Prof. Dr. Zhimin He

2003-2005 Ph.D. at Department of Medical Biochemistry and Biophysics, Umea University, Sweden.

Structural and functional studies of plasminogen activator inhibitor 2 (PAI-2), a serine protease inhibitor.

Supervisor: Prof. Dr. Tor Ny and Prof. Dr. Malgorzata Wilczynska

(Interrupted. To stay with my wife, I transferred to Zurich)

2005-2008 Ph.D. at Institute of Organic Chemistry, Zurich University, Switzerland.

The structural studies of larger N-terminal domain of Y receptors.

Supervisor: Prof. Dr. Oliver Zerbe

Publications

1. **Zou Chao**, Sowmini Kumaran, Zerbe Oliver*. "Biosynthesis and NMR studies of a double transmembrane domain from Y4 receptor". *Journal of Biomolecular NMR*, **2008**: 42, 257-269.
2. **Zou Chao**, Sowmini Kumaran, Reto Walser, Zerbe Oliver*. "Properties of N-terminal domain from Y receptors probed by NMR spectroscopy". *Journal of Peptide Science*, accepted.
3. **Zou Chao**, Sowmini Kumaran, Stefan Markovic, Reto Walser, Zerbe Oliver*. "Studies of the structure of the N-terminal domain from Y4 receptor, a G-protein coupled receptor, and its interaction with hormones from NPY family ". *ChemBioChem*, **2008**: 9, 2276-2284.
4. Zerbe Oliver*, Neumoin Alexey, Mares Jiri, Walser Reto, **Zou Chao**. "Recognition of neurohormones of NPY family by their receptors". *Journal of Receptor Signal Transduction Research*, **2006**: 26, 487-504.

Conferences

1. Collaboration with Russian Academy of Science for nano-disc project, Moscow, Russia. September to October, 2008.
2. Swiss Chemistry Society Fall Meeting 2007 held in Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland on September 12th, 2007. (Oral Presentation)
3. XXXII European Conference on Hormones and Cell Regulation GPCR Complexes and GPCR Complexity held in Mont Ste., France on September 13th to 16th, 2007. (Poster)
4. The 2nd Dorothy Crowfoot Hodgkin Meeting held in Zurich University, Zurich, Switzerland on April 27, 2007. (Poster)
5. Swiss Chemistry Society Fall Meeting 2006 held in Zurich University, Zurich, Switzerland on October 13, 2006. (Poster)

Acknowledgements

In the last part of my thesis I wish to express my sincere thanks to the people who help me make this thesis done.

It is impossible to overstate my gratitude to my supervisor Professor Dr. Oliver Zerbe, who not only took me in this group such that I can stay with my wife but also gave me tremendous help with science and more importantly with my mental situation. Ph.D. is a long and tough journey, several times I was almost beaten down and it is his enthusiasm and encouragement that helped me through.

I am really blessed to have such dear colleagues: Alexey Neumoin, a cool Russian but always willing to help so long as you dare make the first step; Jiri Mares, a computer-coupled Czech and he can solve almost all the computer-related problems; Sowmini Kumaran, a lovely Indian and the most warm-hearted person I have ever met; Reto Walser, a really excellent Swiss and his ability to make perfect job in both NMR and biochemical parts makes me jealous.

I would like to thank diploma student Stefan Markovic and post-doc Sivakami Chandrasekaran who had contributed a lot to my project. Though we have not worked together for too long time, it can not stop us from becoming friends who can open up to each other.

I am grateful to Professor Dr. John Robinson, Dr. Katja Zerbe and his group members who have provided the support for solid-phase peptide synthesis and shared their equipments and thoughtful discussions.

I also appreciate the support that our technicians Simon Jurt and Nadja Bross have offered during the daily service work. Of course special thanks should also be given to Dr. Laurent Bigler and Urs Stalder from Mass Spectroscopy service, since they have killed much time to handle many of my naughty proteins.

I am indebted to my parents, Zou Guishan and Zhu Shufan, who gave me life, raised me, taught me and love me no matter what happens. I also owe my thanks to my younger brother and his family, Zou Jie, Leng Bing, Cui Jiarui and Zou Jinxiao.

In the end I would like to thank my lovely wife and the unborn baby. In a foreign country, your accompany takes away the loneliness, your smile adds vigor into my life and your encouragement promotes me to keep forward. Marrying you is the best thing that ever happened in the past 5 years. I am sure raising the baby together with you will be the most exciting thing in the rest of my life.